Affinity Partitioning and Purification of Plasmid DNA in Aqueous Two-Phase Systems

Hélder Simão da Costa Barbosa

UMinho | 2009

Universidade do Minho
Escola de Ciências

Hélder Simão da Costa Barbosa

Affinity Partitioning and Purification of Plasmid DNA in Aqueous Two-Phase Systems

September 2009

Universidade do Minho
Escola de Ciências

Hélder Simão da Costa Barbosa

Affinity Partitioning and Purification of Plasmid DNA in Aqueous Two-Phase Systems

Thesis for Doctoral Degree in Sciences

Elaborated under the Supervision of:
Professor João Carlos Marcos
Department of Chemistry, University of Minho, Portugal

Professor Nigel K. H. Slater
Department of Chemical Engineering and Biotechnology,
University of Cambridge, United Kingdom

September 2009

Universidade do Minho, ___/___/_____

Assinatura: _____

The work presented in this thesis was done mainly within the Center of Chemistry, School of Sciences, University of Minho, Portugal and in the Department of Chemical Engineering and Biotechnology, University of Cambridge, United Kingdom. Part of the work was also done in the Department of Pharmaceutics, School of Pharmacy, University of London, United Kingdom. The financial support was given by FCT, the Portuguese Foundation for Science and Technology (PhD grant reference: SFRH/BD/16296/2004/XS49).

Acknowledgements

The work reported in this thesis would not have been possible without the mentorship of many people. Firstly, I would like to express my deepest gratitude to Professor João Carlos Marcos from University of Minho for having taken me as his PhD student and for his wise guidance and encouragement over the course of this project. All this work would not have been possible without him.

I gratefully acknowledge Professor Nigel Slater from the University of Cambridge (UK) for having given me the opportunity to join his research group during a large period of my studies. I really admire his enthusiasm towards research, which motivated me to achieve my best. I was always amazed by the clearness of his ideas and scientific knowledge.

I also extend my sincere appreciation to Professor Steve Brocchini from the University of London (UK) for also having welcomed me in his group and for his great assistance through all the PEGylation work. I have learnt a lot from him and I truly admire his awareness for the importance of the details in research.

Thanks are due to Doctor Anna Hine from Aston University (UK) for kindly having provided me with the plasmids vectors used in this work and also for the kind support and assistance throughout the research work.

I would also like to express my gratitude to all the lab members (and they were many) in all the three research groups for their valuable help and friendship. Having spent time with them in the lab (and outside of it) was an immense pleasure and a great joy.

I would like to thank also all the support of my close friends and family. Thanks for always being there and for all the encouragement.

Finally, a special thanks to my dear mother, "Vita". This thesis is dedicated to you.

To all, MUITO OBRIGADO!

To my mom

Abstract

Plasmid DNA (pDNA) vectors are being developed as preventive or therapeutic DNA medicines for the treatment of parasitic, bacterial and viral diseases and also for other indications, such as the treatment of devastating inherited diseases. This has been reflected in a demand for efficient and cost-effective large-scale processes for the production of pharmaceutical-grade pDNA.

The utilisation of aqueous two-phase systems (ATPS) for the extraction of pDNA has several advantages over the conventional chromatographic processes; namely, easier scaling-up, operation simplicity, higher capacity, and the possibility to combine purification, concentration and clarification into one unit operation. However, the main drawback of ATPS is the low selectivity for the target pDNA in complex *Escherichia coli* cell lysates. In this thesis two protein-based affinity systems were investigated in order to enhance selectivity of the ATPS by exploring sequence-specific recognition of target pDNA:

1) **GST-ZnF/pTS**: a fusion protein, comprised of Glutathione-S-Transferase (GST) and a zinc finger transcription factor (ZnF) that binds to 5´-GGGGCGGCT-3´ sequence within pTS plasmid.

2) **LacI-His$_6$-GFP/pUC19**: a *lac* repressor protein (LacI) fused to a His$_6$-tag and to a Green Fluorescence Protein (GFP) that displays affinity for two *lac* operator sequences (*lac*O$_1$/*lac*O$_3$) within pUC19 plasmid.

Initially, polyethylene glycol (PEG) - dextran (DEX) systems were selected for the favourable partitioning of the pDNA impurities to one of the phases, i.e. dextran. In this way the affinity ligand would steer the pDNA to a phase with less impurities, so increasing its purity and the selectivity of these systems. The PEG 600 – DEX 40 system was selected with all major pDNA impurities accumulated predominantly in the bottom DEX phase (RNA >99 % and protein >80 %) with total pDNA recoveries >80 %.

In order to efficiently steer the affinity ligand into the top phase, the GST-ZnF fusion protein was covalently bound to PEG polymer (PEGylation). Initially, four

different PEGylation reagents were optimized for conjugation to the free GST. Thiol selective PEGylation reactions were completed after 1h under reducing conditions. By contrast, amine selective reagents required much longer reaction times (24h) or high PEG equivalents, in order to yield more than 80% conjugation. Optimized reactions on free GST were then applied to GST-ZnF fusion protein resulting in similar PEGylation yields. The PEG-GST-ZnF ligand was then used for the complete isolation of pre-purified pTS plasmid from bottom phase to the top phase in PEG 600 – DEX 40 systems. The PEGylated ligand accumulated 97.5 % in the top phase, while 96.5 % of the native GST-ZnF protein and 99.9 % pDNA accumulated on the bottom. Although the proof-of-concept for a pDNA affinity extraction process was demonstrated, no further assays were carried out using bacterial cell lysates due to the tight protein–DNA binding which make the elution of the pDNA impractical.

An alternative approach to attach PEG to the affinity ligand was tested with an immobilized metal–ion affinity PEG (PEG-IDA-Cu^{2+}) which binds to the his-tag accessible on the surface of LacI-His_6-GFP protein. The affinity extraction of pUC19 plasmid from a bacterial cell lysate in PEG 600 – DEX 40 ATPS containing 0.273 nmol of LacI fusion protein and 0.14 % (w/w) functionalised PEG, resulted in more than 72 % of the plasmid DNA partition to the top phase, without RNA or genomic DNA impurities. Although pDNA elution from LacI binding complex proved to be difficult, the method described removed more than 75 % of the protein yielding ~27 % of pCU19 pDNA. This represents about 7.4 μg of pDNA extracted per mL of desalted lysate which is about 9 fold improvement to what was previously obtained using LacI-His_6-GFP immobilized into a chromatographic column.

To conclude, it is envisaged in this thesis that the knowledge obtained using pDNA affinity ligands would contribute for the development of a highly selective plasmid DNA manufacturing process fulfilling the stringent specifications and demands for gene therapy or DNA vaccines applications.

Resumo

Vectores de DNA plasmídeo (pDNA) estão a ser desenvolvidos para a prevenção e tratamento de doenças do foro viral, bacteriano e parasitário, bem como, para o tratamento de doenças crónicas hereditárias. Nesse sentido, existe a necessidade crescente de processos eficientes e económicos de produção industrial de pDNA com qualidade farmacêutica.

A utilização de sistemas de duas-fases aquosas (SDFA) para a purificação de pDNA tem várias vantagens comparativamente aos processos cromatográficos convencionais, nomeadamente a facilidade de aumento de escala, maior capacidade e a possibilidade de combinar numa única operação os processos de purificação, concentração e clarificação. No entanto, a maior desvantagem dos SDFA deve-se à sua baixa selectividade para os vectores pDNA em lisados celulares de *Escherichia coli*. Nesta tese, foram investigados dois sistemas de afinidade com o objectivo de aumentar a selectividade dos SDFA utilizando ligandos proteicos com especificidade de ligação ao pDNA:

1) **GST-ZnF/pTS**: proteína fusão, composta pela proteína Glutationa-S-Transferase (GST) e um factor de transcrição dedo do zinco (ZnF), que se liga à sequência 5´-GGGGCGGCT-3´ do plasmídeo designado por pTS.
2) **LacI-His$_6$-GFP/pUC19**: proteína fusão, constituída pela proteína repressora do operão *lac* (LacI) fundida com 6 resíduos de aminoácidos de histidina e com a proteína verde fluorescente (GFP), que se liga a duas sequências *lac* operão (*lac*O$_1$/*lac*O$_3$) do plasmídeo pUC19.

Inicialmente, SDFA com polietilenoglicol (PEG) e dextrano (DEX) foram seleccionados de acordo com a acumulação preferencial na fase de DEX (ou fase inferior) dos contaminantes principais do pDNA presentes nos lisados. Deste modo, os ligandos de afinidade permitem extrair o pDNA para a fase com menos contaminantes - fase de PEG - aumentando assim o grau de purificação e selectividade dos SDFA. O sistema PEG 600 – DEX 40 foi seleccionado para o efeito, obtendo-se a acumulação do RNA (>99 %) e das proteínas (>80 %) na fase inferior e com recuperação superior a 80 % do pDNA.

Para facilitar a partição do ligando para a fase superior, a proteína GST-ZnF foi ligada covalentemente ao polímero PEG (PEGuilação). Inicialmente, 4 reacções de conjugação foram optimizadas usando a proteína GST livre. A PEGuilação dos grupos tiol foi total ao fim de 1 h em condições redutoras. Conjugações superiores a 80 % com os grupos amina foram obtidas ao fim de 24 h ou usando uma maior razão reagente/ligando. As reacções optimizadas foram de seguida aplicadas à proteína fusão GST-ZnF, resultando em rendimentos semelhantes. O ligando PEG-GST-ZnF foi posteriormente usado para transferir totalmente o plasmídeo pTS da fase inferior para a fase superior em sistemas PEG 600 – DEX 40. O ligando PEGuilado acumulou-se 97,5 % na fase superior, enquanto que a proteína GST-ZnF nativa e o plasmídeo pTS, acumularam-se 96,5 % e 99,9 % na fase de DEX, respectivamente. Embora o "*proof-of-concept*" para um processo de extracção de pDNA por afinidade tenha sido estabelecido, o processo não foi testado com lisados bacterianos dado que a ligação GST-ZnF/pDNA é extremamente forte e a sua eluição impraticável.

Foi ainda desenvolvido um processo alternativo para ligar PEG ao ligando, usando PEG funcionalizado com iões metálicos (PEG-IDA-Cu^{2+}) com afinidade para as histidinas à superfície da proteína LacI-His_6-GFP. A extracção do plasmídeo pUC19 de lisados celulares em SDFA foi realizada com a presença de 0,273 nmol do ligando LacI-His_6-GFP e 0,14 % (w/w) de PEG funcionalizado. Cerca de 72 % do pUC19 acumulou-se na fase superior, sem a co-purificação de RNA ou DNA genómico. Embora a eluição do pDNA se tenha provado difícil, a metodologia desenvolvida permitiu remover 75 % da proteína total e obter um rendimento final de 27 % de pUC19 (7,4 μg pUC19 por mL lisado). Estes resultados correspondem a uma capacidade 9 vezes superior ao método cromatográfico descrito na literatura com o mesmo ligando.

Em suma, este trabalho resultou no aumento de conhecimento na utilização de ligandos de afinidade em SDFA que poderá contribuir para o futuro desenvolvimento de processos industriais com alta especificidade para o pDNA, satisfazendo as especificações e procura de pDNA para terapia genética e vacinas de DNA.

Table of Contents

Abbreviations List

A - *Adenine (organic base)*

AEX - *Anion exchange chromatography*

ATPS - *Aqueous two-phase systems*

BCA - *Bicinchoninic acid*

bp - *Base pairs*

BP - *Bottom phase*

BSA - *Bovine serum albumin*

C - *Cytosine (organic base)*

Cu(II) - *Copper ion*

Da - *Dalton (Dalton x 1.66×10^{-24}= grams)*

DEX - *Dextran*

DNA - *Deoxyribonucleic acid*

dsDNA - *Double strand DNA*

EDTA - *Ethylene diaminetetraacetic acid*

E. coli - *Escherichia coli*

FDA - *Food and Drug Administration*

G - *Guanidine (organic base)*

gDNA - *Genomic DNA*

GST - *Glutathione-S-Transferase*

GST-ZnF - *Glutathione-S-Transferase-Zinc finger fusion protein*

IDA - *Iminodiacetic acid*

HIC - *Hydrophobic interaction chromatography*

K_D - *Dissociation constant*

kDa - *Kilo Daltons*

LacI - *Lactose repressor protein*

LacO - *Lactose operon*

LacO1 - *First lacO binding site*

LacO2 - *Second lacO binding site*

LacO3 - *Third lacO binding site*

OD - *Optical densitiy*

PEG - *Polyethylene glycol*

PEG-ALD - *mPEG-butyraldehyde*

PEG-MAL - *mPEG-maleimide*

PEG-MS - *mPEG-monosulfone*

PEG-SPA - *mPEG-succinimidyl propionate*

PBS - *Phosphate buffer saline*

pDNA - *Plasmid DNA*

pI - *Isoelectric poin*

pUC19 - *pUC19 Plasmid (plasmid that affinity binds to LacI)*

pTS - *pTS Plasmid (plasmid that affinity binds to ZnF)*

RNA - *Ribonucleic acid*

SDS - *Sodium dodecyl sulphate*

SDS-PAGE - *SDS-Polyacrylamide gel electrophoresis*

SEC - *Size exclusion chromatography*

ssDNA - *Single stranded DNA*

scDNA - *Supercoiled DNA*

T - *thymine (organic base)*

TP - *Top phase*

ZnF - *Zinc finger (i.e. the Zinc finger motif of the GST-ZnF fusiom protein)*

Objectives and Thesis Layout

The major aim of this project was to investigate and discuss whether protein-based affinity ligands could be exploited for the large-scale purification of plasmid DNA in aqueous two-phase Systems (ATPS). The affinity extraction of pDNA in ATPS has generated interest, since it has high potential to improve the low selectivity of these systems combined with high pDNA capacities, faster processing times, and potentially more cost-effective compared to the traditional chromatographic methods. In the quest to accomplish the general objective, this thesis is organized in 5 individual chapters each aiming to fulfill a specific goal:

Chapter 1 presents a general introduction that reviews the central biochemical challenges on plasmid DNA production and overview the current advantages and disadvantages of the existent pDNA purification methods.

The **Chapter 2** focus on the selection of suitable ATPS for the affinity extraction of pDNA by studying the partitioning of total DNA, total protein and total RNA from a bacterial cell alkaline lysate in PEG – dextran systems.

In **Chapter 3** results are presented on the optimization of protein PEGylation reactions and on the utilisation of covalent PEGylated and non-PEGylated GST-ZnF affinity ligand for the affinity partitioning of pDNA bearing the zinc finger recognition site (pTS). **Subchapter 3.1** focus on the optimization of four different PEGylation reactions for either amine or cysteine PEG conjugation on the free GST-tag protein. The **Subchapter 3.2** presents the affinity partitioning of pTS plasmid using the modified and native GST-ZnF fusion ligand in the previously selected systems.

The **Chapter 4** focus on the utilization of the sequence specific DNA binding protein, LacI-His_6-GFP, and a metal affinity PEG-IDA-Cu(II) as a dual affinity mechanism for the pUC19 plasmid extraction from a bacterial cell lysate in ATPS systems.

Finally, **Chapter 5** is devoted to an integrative discussion focused on the main contributions of the present work to the understanding of affinity purification of

plasmid DNA in aqueous two-phase systems. Brief suggestions for future directions to complement our research are presented, together with a few general concluding remarks.

Chapter 1

General Introduction and Literature Review

1.1 Why Purify plasmid DNA

1.1.1 Gene Therapy and DNA Vaccines

The dawn of the modern biotechnological era can be traced back to the discovery of the DNA structure by James Watson and Francis Crick in 1953 [1]. However it was the discovery of basic tools of recombinant DNA, including the DNA ligases and restriction enzymes, that led Paul Berg and colleagues, to the creation of the first recombinant DNA in 1972 [2]. In 1973, Herbert Boyer and Stanley Cohen created the first recombinant DNA organism [3], a genetic engineering breakthrough with implications in the development of whole new therapies approaches for the treatment of a widespread type of diseases in the following decades.

Gene therapy is one of the fields that in recent years has attracted a lot of interest, which is expected to grow more promptly in the future as the whole human genome has been sequenced and is now being annotated for potential therapeutic targets [4; 5]. The high potential of this therapy strategy is owed to the possibility to cure by the introduction of nucleic acids into human cells in order to restore, enhance or cancel a specific element on the genetic repertoire. Unlike other traditional treatments, gene therapy aims to remove the cause of a disease and not to correct the biochemical imbalance, or getting rid/attenuate the disease symptoms as in the traditional treatments [5].

Traditionally, patients with particular genetic disorder are treated by the administration of therapeutic drug, e.g. proteins, that have to be taken frequently in order to tackle the disease. This is because proteins in the blood stream are degraded or eliminated with time and the therapeutic effect is limited to a short period. Gene therapy may circumvent these problems because the gene encoding the therapeutic protein is stably placed inside the cell and thus may result in longer-term expression of the therapeutic proteins [5].

Commonly, the nuclei acid used is the double-stranded DNA (dsDNA) encoding a specific protein but other nuclei acids such as the single stranded

DNA (ssDNA) or antisense of RNA, have also been described [6]. Antisense RNA binds to a target sequence in the host cell and inhibits the expression of a specific gene by blocking either gene expression or its promoters.

Alternatively, these transgenes can encode a target antigen that will elicit both humoral and cellular immune system's response to the specific antigen. This process is designated by DNA vaccination [7]. The genetic transfer vector uses the host cell protein machinery to produce the antigen protein that will eventually emerge on the cell surface as Major Histocompatibility Complex (MHC) molecules, thereby inducing an immune response to the antigen [7; 8].

One of the critical challenges in gene therapy and DNA vaccines is the efficiently and safety delivery of the nucleic acids into the target cells. Viral and non-viral delivery vectors have been developed for this purpose [9]. The use of viral vectors (e.g. retrovirus, adenovirus) have been questioned due to safety and regulatory concerns because of their toxicity and strong immune response after repeated administration [6]. In addition, the possibility to activate oncogenes or suppress tumor-suppressor genes have raised apprehension in the application of this vectors [9]. In fact, the viral vectors are responsible for some of biggest gene therapy setbacks as deaths on clinical trials have been reported [10].

The development of non-viral vectors has thus emerged as a safer alternative to viral vectors and led to the study of naked plasmid deoxyribonucleic acid (pDNA) or encapsulated/complexed pDNA as vector for transferring prophylactic and/or therapeutic genes. It is generally simpler to develop compared to the viral counterparts, and has the advantage of being free from most of safety concerns associated to the virus [11].

However, the main disadvantage of non-viral vectors has been the insufficient level of production of protein/antigen by the host cell, thus triggering generally weak immune response. In addition, short duration of the expression has also been described [12]. At the present, virus vectors are the most efficient gene delivery systems and, regardless of the safety concerns, they are still dominant and widely applied in clinical trials. More than 66 % clinical trials worldwide

are currently being performed using attenuated virus, such as the adenovirus, retrovirus or vaccinia virus; whereas naked / plasmid DNA represents only 17.9 % (246 ongoing clinical trials) [13].

Despite the low transfection efficiencies, plasmids do not seem to activate anti-DNA antibody secreting B lymphocytes, thus they are usually recognized as having better safety profiles and improved stability compared to the viral vectors [8]. Researchers are hoping to overcome the non-viral vectors disadvantages by co-expressing enhancing agents within the cells and by mimicking live virus infections [14].

Other area of concern is the potential integration of the pDNA into the host cell genome. Although, the integration into the host genome has not been observed after 18 weeks of pDNA administration, concerns of this possibility remains a main issue [15]. There are stringent regulatory controls in order to prevent these problems to happen and thus a pDNA product has to comply with high levels of safety, quality and efficacy in order to be licensed [6; 16]. Nevertheless, it is believed that all the hazards of the pDNA side effects will be overcome in a near future. The potential for the improvement of current disease treatments or vaccination are the most promising attributes of the pDNA-based therapeutics, which eventually will transform the future healthcare.

1.1.2 Current demands for plasmids

In gene therapy, the therapeutic effects of plasmids are due to the DNA sequence that they carry, which, generally, encodes a protein that will induce the desirable therapeutic effect. The first successful direct transfection of animal cells with plasmid DNA *in vivo* was reported in 1990 by Wolff *et al.* [17] in a mouse. Since that time, the gene therapy field has developed very fast and now, there are already widespread gene-based indications for human use, which are currently in clinical trial phase. As last updated in March 2009, the foremost clinical trials worldwide addressed by gene therapy (n= 1,537) are for cancer diseases (64.6 %), cardiovascular (8.9 %), monogenetic (8.1 %) and infectious

diseases (7.9 %) such as HIV, malaria, and tuberculosis [13]. Naked / plasmid DNA represents about 18 % of the total trials and this represent clearly a rising demand, as figures in 2000, were only about 4 %. [18]

The increasing usage of pDNA vectors in gene therapy/DNA vaccination trials results in an urgent need for development of large-scale and cost-effective manufacturing process of pDNA in order to meet those demands. Since pDNA has commonly low transfection efficiencies, it is also expected that larger doses of the DNA-based product are needed, or repeated administrations of the pDNA will be necessary in order to achieve successfully gene transfer [6; 12]. An average dose size will differ between therapies and each disease is likely to have specific needs. It is expected that typical doses size may vary between 0.1-5 mg of pDNA, but full treatments may require larger quantities of plasmid DNA [19]. Thus, great amounts of pDNA have to be produced in order to be sufficient to supply for world requirements [6; 20].

Some companies are already preparing and launching new facilities for the large-scale pDNA production. Aldevron LLC, for example, is a USA biotech company, which opened a facility in 2007 to manufacture kilogram amounts of pDNA. They claim to be sufficient for the production of up to 2,000,000 DNA vaccine doses assuming that a dose contains 500 µg of pDNA [18]. This clearly shows that really large amounts of the therapeutic vector may be required and this poses several challenges for the biochemical engineers. In addition, when the DNA vaccines prove to be efficient, safer and efficient alternative to the traditional vaccination methods, the demand for these pDNA vectors will increase dramatically.

DNA vaccines will have an important role in the fight against virus pandemics outbreaks as the influenza virus outbreak (H1N1) we are currently suffering [21]. Compared to conventional therapeutic methods, DNA vaccines can be cheaper to produce, more stable at ambient temperature and safer to administrate [22]. Moreover, DNA vaccines can be produced as short as in two weeks, contrasting to 8-9 months of the actual methods [21]. This represents

clearly a huge advantage since a much rapid-response to an epidemic outbreak can be achieved. Kieny *et al*. [23] have calculated that for the prevention of epidemics, like the H1N1 virus, it will require the vaccination of 60 % of the world population (6.8 billion, as last revised 2008 [24]) and it has been estimated that existing production methods are able to supply only 350 million doses per year [18].

In Portugal, the government recently placed an initial order of 6 million doses of the H1N1 influenza virus vaccines (costing €45 Million) for the initial vaccination of about 30 % of the Portuguese population [25]. Orders of this magnitude have been placed all over the world in order to stop a potential spread of the virus between populations. In fact, several countries have placed orders higher that their entire populations, in the case repetitive doses are needed; Canada [26], United Kingdom [27], Israel [28] are amongst them. Ultimately, the replacing of the current vaccines by the DNA vaccines will make these a leading "drug" in the new generation of blockbuster therapeutics, since the demand for these vectors will be huge.

1.1.3 Current gene-based products / market size

The first gene therapy product is currently being commercialised only in China since 2004 with the trade name Gendicine™ (Sibione GeneTech) for the treatment of head and neck squamous cell cancer [29]. The second licensed gene-based product has been also approved only for China market in 2006, under trade name of Oncorine™ (Sunway Biotech), for the treatment of the same cancer. Cerepro® from the Anglo-Finnish company Ark Therapeutics Group PLC, will be probably the first gene-based product to be commercialized outside china, as European Medical Agency (EMEA) is currently reviewing it for marketing authorisation in European Union [29]. All the three products described above are virus-based drugs for human use.

Nonetheless, as March 2009, there are currently 246 ongoing gene therapy trials worldwide that utilize plasmid DNA, of which 2 are already in clinical

trial phase III [13; 30]. Thus, it is expected these products to reach the market in 2-5 years if it gets approval for commercialization.

However, four pDNA-based therapeutics have been already licensed in the veterinary environment. The world's first licensed DNA vaccine was attributed in 2005 for the immunization of horses against West Nile Virus (Ft Dodge Animal Health). Later it was also approved a pDNA vaccine for necrosis virus immunization in salmon (Novartis), for the melanoma cancer treatment in dogs (Merial) and for the growth hormone therapy in Pigs (VGX Animal Health). As the pDNA effectiveness in animals is been demonstrated and well established, soon it is also expected the same successes for such pDNA applications in humans [7; 30].

The market size for gene therapy/DNA vaccines in humans is difficult to estimate, as the only two gene-based products were approved to the Chinese market and the true potential of the DNA-based drugs are still under evaluation. Nonetheless, the exponential increase in the number of patents issued in the area of DNA vaccines indicates that DNA-based drugs have generated a huge interest of many, demonstrating that pDNA could be a promising therapy technology in the future. In 1990 there were only 2 patents contained the term "DNA vaccine" whereas by time this thesis is written there are 1223 issued patents index by the United States Patent and Trademark Office [18; 31].

The revenue predication to be generated from sales of these products is in the order billions of dollars worldwide [6]. However, since the human genome seems to encode 20,000-25,000 protein-coding genes [4], and their therapeutic value being asserted, the potential value of gene therapy could be as high as trillions of dollars globally. Therapeutic plasmid DNA is an example of a new generation of high-value low-volume medicines.

1.2 Central biochemical engineering issues

There are three main stages on the manufacturing of plasmid DNA, Figure 1. The first stage is the construction of the vector and the transformation to a suitable host strain and is designated as upstream processing. The second stage is the cultivation / fermentation of pDNA host cells in order to amplify the pDNA. The third stage consists of the purification and formulation of plasmid DNA vectors and is designed as pDNA downstream processing. All the three stages are interconnected; hence each former step may have an impact on the efficiency of the following steps. In all these steps there are several critical challenges and difficulties, which the biochemical engineers have to address in order to produce pharmaceutical grade pDNA.

This introduction does not set out to cover all these initial aspects of the plasmids manufacturing, consequently only a general perspective of some important aspects on the upstream processing of pDNA will be given. Reviews have been published elsewhere by Williams *et al.* [30] and by Tejeda-Mansir *et al.* [32] covering some important aspects of the pDNA upstream processing, which are not covered here but cannot be neglect for efficient plasmid DNA production.

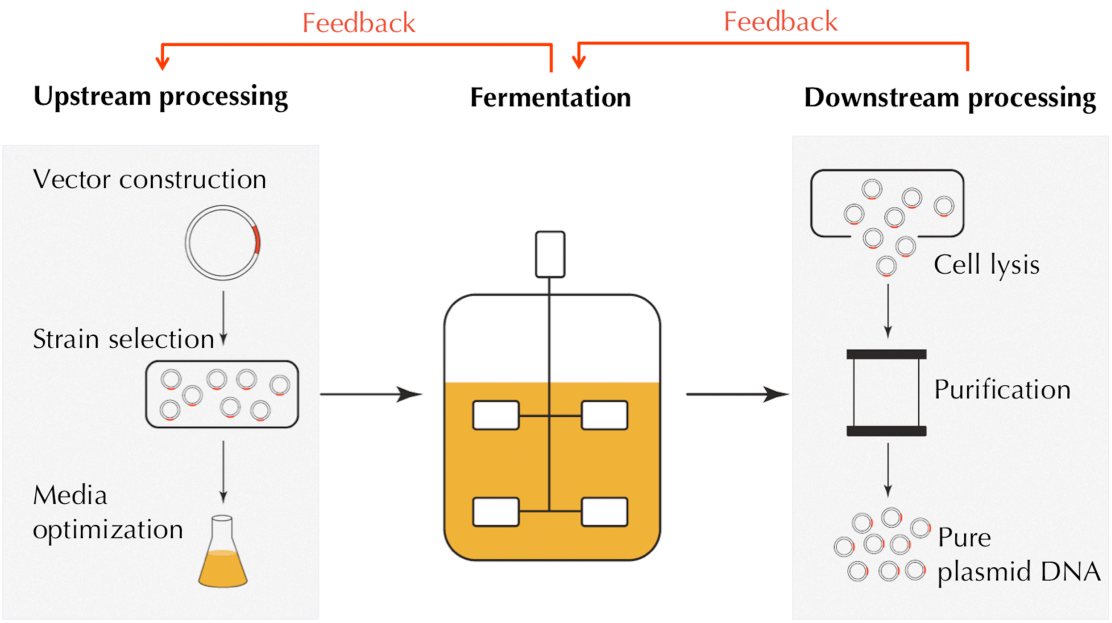


Figure 1 – Overall process for the plasmid DNA manufacturing. Adapted from [6].

1.2.1 Vector selection and design

The judicious selection of the plasmid vector is determinant for the production of high pDNA yields and for pDNA stability (i.e., the tendency of bacteria to completely loose the plasmid during cultivation) but also to increase transfection efficiencies and safety [30]. A careful design of the backbone vector is essential in order to fulfil with all the regulatory license guidelines.

For efficient gene expression, pDNA must contain the desirable gene but also other important elements such as intron(s), eukaryotic promoter, polyadenylation sequences and, possibly an operon region. The majority of the plasmids vectors currently in used on DNA vaccines trials are derived from pBR322 or pUC plasmids, and use high copy derivatives of the pMB1 origin (closely related to the ColE1 origin)[14; 30]. The plasmid copy number in each host bacterial cell is important for greater yields but also to facilitate its downstream processing since the ratio pDNA to gDNA, RNA and cellular proteins (pDNA impurities) is increased. However high copy plasmid, such as the ones with pUC origin of replication, tend to reduce the growth rate, which over a period of time, may result in cells that has lower pDNA content [14; 30]. This phenomenon results of lack of active partitioning functions, as plasmids are not precisely distributed between cells. Consequently, plasmid-free cells arise as the propagation time increases [30].

In order to increase the number of cells carrying the therapeutic gene, often the DNA vector carries a gene that confers resistance to an antibiotic. This gene is important for the selective growth of the cells with the vector to dominate. The common antibiotic resistance gene is *KanR* which confers resistance to kanamicin antibiotic [33]. Ampicillin resistance gene (*Amp*) is discouraged since some patients may have hyperactivity to β-lactam antibiotics [30]. Nevertheless, *Amp* is commonly used on vectors for non-therapeutic use. To ensure safety, regulatory agencies recommend elimination of antibiotic resistance markers from therapeutic pDNA vectors.

Interestingly, the selection via antibiotic resistance does not also necessarily

lead to the production of high titer of plasmid DNA in the bacteria due to segregation instability during cell division [34]. Thus, several other antibiotic-free selections for the pDNA production have been developed, such as a RNA selective marker (as shown in a model plasmid vector represented in Figure 2), amino acid auxotrophy complementation [35], strains lacking the essential gene initiation factor 1 [36] or *Lac* operator [14].

A representative example of RNA selective marker was developed by Luke *et al.* [37]. The authors have incorporate and express a 150 bp RNA-OUT antisense RNA that represses expression of a chromosomally integrated constitutively expressed counter-selectable marker (sacB), allowing plasmid selection on sucrose. This approach resulted in pDNA yield as high as 1 g/L.

Antibiotic-free selection is a safest and most recommended alternative since the presence of the gene marker is discouraged due to concerns regarding the dissemination of antibiotic resistance genes to patient's enteric bacteria [30].

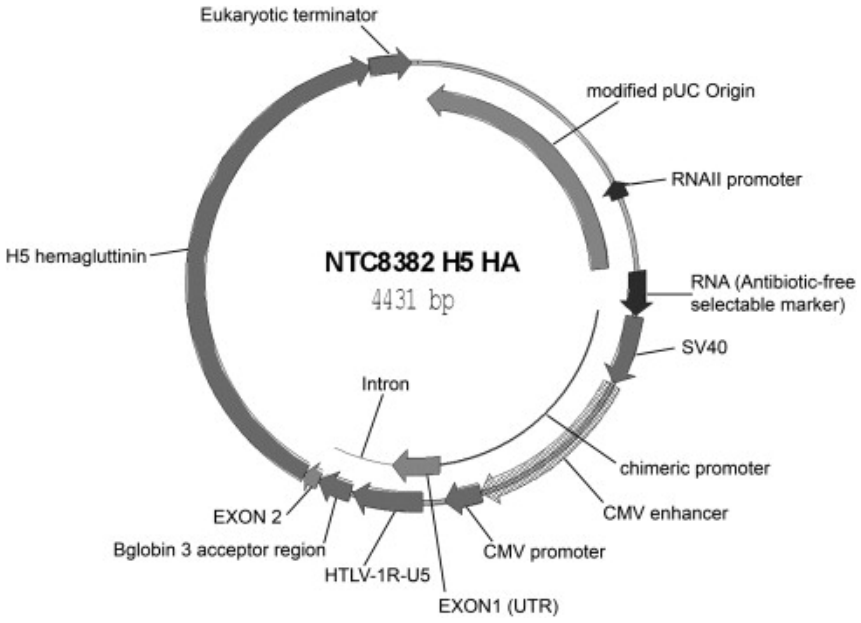


Figure 2 - Antibiotic free bird flu DNA vaccine vector. Vector specific features such as the CMV-HTLV-1 R-U5 derived mRNA leader (UTR) and RNA selective marker are indicated [30].

Commonly, DNA vectors have also the human cytomegalovirus/immediate-early or CMV-chicken-β actin promoters because it provides highest levels of gene expression in a variety of cell types and tissues. These promoters usually lack any DNA sequences that are not essential for their function, therefore minimizing the possible homology with the host [33; 38].

Examples of some backbone vectors currently in use in some clinical trials are the pVAX1 (Invitrogen), V1JNS (Merck), pCMVKm2 (Chiron) but many others vectors, having different features, are also been currently in use [30].

In summary, the design of pDNA vector is thus of paramount importance for the overall success of DNA vaccines production and applications. Varied aspects on the plasmid design cannot be given negligence since the impact on the following upstream and downstream processing can be significant, as well as, it may have an impact on the safety for human administration.

1.2.2 Fermentation

Plasmids are traditionally propagated in *Escherichia coli* (*E. coli*) K12 strains such as DH5α, DH1, JM108 [30]. Fermentation strategies combines the selection of fermentation conditions with the host organism and construct design to maximize both volumetric yield (mg/L) and specific yield (mg/OD_{600}/L) of high quality supercoiled form plasmid DNA.

Plasmid copy number increases as the specific growth rate decreases [38]. Thus, the selection of cell growth media and the fermentation mode has to be cautiously performed. For example, some studies have found that the pDNA content increases by tenfold as the carbon/nitrogen ratio increases from approx. 0.2 to 2.78 [39]. Higher pDNA yields were also obtained using amino acid starvation or with the addition of chloroamphenicol to the growth media. Furthermore, the use of a heat shock (shift from 30 to 42 ºC) have also shown to increase the plasmid yield, nonetheless, at large scale, heat shock may be

difficult to apply [14]. The use of chloroamphenicol may also pose some regulatory concerns and thus may not be a suitable option.

Fed batch cultivation, employing an exponential feed regime, is the most efficient mode for controlling specific growth rate (growth rate < μmax) and thus enrichment of pDNA in the biomass contrasting with batch mode. Both methods have been successfully employed for plasmid production with *E. coli* systems, but the fed-batch mode is the most dominant approach in several pDNA production operations [18].

The fermentation optimization process has three main heuristics that are always desirable to be achieved. The first is the cell culture mass per liter of fermentation. Evidently, higher cell density yields higher plasmids per unit volume of culture; Secondly, is the plasmid copy number per each cell host, which also leads to high specific yields; and third the plasmid quality, i.e., supercoiled pDNA content (see Section 1.2.4), to meet the regulatory demands for approval.

Changes in growth conditions including nutrients concentrations, temperature, pH, oxygenation and growth phase can affect the pDNA quality, so these parameters should be carefully monitored. The production of plasmid DNA at high titer depends on the selection of suitable strategies for the control and regulation of bacteria growth and nutrient supply, but, as shown above, the optimization process starts at the plasmid design stage.

Currently, a plasmid titer as high as 2.2 g/L is the benchmark [30]. At this productivity, 1 kg of plasmid in a 500 L fermentor could be generated. Yet, it should be noted again, the titer is not the only important fermentation criterion, as maintaining high pDNA quality is also are a major issue. In fact, the major problems in the pDNA manufacturing at the present resides mainly on the improvement of the bottleneck cell lysis and in the downstream processing of pDNA to solve this mismatch [19].

Representative examples of patents using fed-batch fermentation processes are issue to Merck [40] and to Boehringer Ingelheim using the *E. coli* JM108

strain [41].

1.2.3 pDNA Impurities

One of the major concerns on the production of clinical grade pDNA is the assurance that the gene-based products have stringent impurities clearance. The high standard removal of all these impurities is essential for product approval as therapeutic drug, which represents one of the major concerns on the scaling-up production of plasmids. The regulatory agencies limits the impurities contents in each DNA formulation as summarized in Table 1.

Table 1 – Impurities specifications and content in bacterial cell lysates. Adapted from [42]

Content of bacterial cell lysates	Final Product		
	Issue in focus	**Range of acceptance**	**Determined by**
55 % Proteins	Protein	<3 µg/mg pDNA	BCA test
3 % host gDNA	gDNA	<2 µg/mg pDNA	TaqMan-PCR
3 % LPS	Endotoxins	<10 E.U./mg pDNA	LAL test
>3 % Plasmid	scDNA over oc form	>97 %	CGE
21 % RNA*	RNA	<0.2 µg/mg pDNA	Analytical HPLC
15 % Others			

* 53 % 23S rRNA; 27 % 16S rRNA; 2 % 5S rRNA; 14 % tRNA; 4 % mRNA

The required pDNA product specifications shown above impose several problems on the downstream processing of pDNA, resulted by the relative low amounts of the target pDNA molecule in the *E. coli* bacterial cell lysates. Plasmid DNA accounts for less than 3 % of total content of a typical cell lysate, which increases the difficulty of its extraction. In addition, the pDNA is structurally related with other impurities such as RNA and with the genomic DNA. Both these impurities are in higher concentrations in the lysate and share some physicochemical properties with pDNA, such as, size and charge.

Non-fragmented gDNA is a large molecule, which has a molecular weight of 2700 MDa, whereas pDNA may vary between 2-20 MDa, which is significantly smaller [43]. However, inefficient alkaline lysis and the shear effects during

preparation of the lysate may fragment the gDNA into smaller portions with similar sizes and charge densities to pDNA, making the removal of genomic DNA particularly challenging. It is being observed that the size of pDNA and gDNA are on the major constraint in the physical operations due to their susceptibility to fluid mechanical forces. This represents one of the most critical bottlenecks in the large scale purification of plasmid DNA in all the spectrum of pDNA downstream processing techniques available [43].

A non-nucleic acid impurity consists mainly of endotoxins and proteins. Bacterial endotoxins are negatively charged (net charge) lipopolysaccharide elements of the outer membrane of gram-negative bacteria. When cells are broken down during the cell lysis, endotoxin globules of widely varying sizes can appear. In humans, endotoxins can cause responses ranging form slight effects of septicaemia to severe effects such as intravascular coagulation [44]. Thus, the removal of endotoxins from a DNA formulation is also crucial but since it also exhibits similar sizes to plasmids (and charge) they are considered difficult to remove.

Endotoxins have affinity to almost every material and/or solvent used in the purification process, and it has been observed that endotoxins concentration decreases rapidly when a chain of purification systems is employed. Thus, at the end of multi-stage purification process the endotoxins are usually in sufficiently reduced concentrations to meet the regulatory demands without the need for specific endotoxin-removal operation [45; 46]. Nonetheless, affinity matrices such as polymyxin B [47], thiophilic aromatic chromatography [48] and HIC (in some extent) [49] have been optimized for endotoxin removal.

On the other hand, proteins impurities as low as 1ng per dose has been known to give rise to adverse reactions in humans, although the immunogenic response is highly dependent on specific proteins and recipients [50]. Therefore, the efficient protein removal during a pDNA extraction process is also required by the regulatory agencies, which impose a maximum limit of 3 μg/mg pDNA. As seen in Table 2, proteins are physiochemical less similar with pDNA molecules,

and generally, the efficient protein removal does not impose a lot of hurdles on the plasmid downstream processing, since selective conditions for pDNA can be found more easily.

Table 2 – Comparison of the physico-chemical properties of proteins and plasmids. Adapted from [51].

Characteristics	Biomolecule	
	Protein	**Plasmid DNA**
Building blocks	amino acids	nucleotides
Molecular mass (kDa)	10^3 - 10^5	10^6 - $10^{7a)}$
Stokes radius	< 5 nm	100 > 300 nm
Charge	depending on pI/pH	negative
Diffusion coefficient	medium-high	low
Viscosity	low	high
Shear force sensitivity	low	high

a) The larger size of the plasmids strongly influences all other properties.

1.2.4 Physico-chemical properties of plasmid DNA

In addition to the efficient removal of pDNA impurities, the plasmid DNA formulations have to be in high quality form, i.e. 97 % of the plasmid DNA should be in the supercoiled form (Figure 3) and the content should be 95 % (w/w) of total nucleic acid [6; 42].

Plasmids are circular DNA molecules and there are found in bacteria independent of the chromosomal DNA. The pDNA can exist in three tertiary structures: supercoiled, open circled and linear. As seen in Figure 3, the plasmid structure denoted as supercoiled is a circular double-strand DNA structure where the DNA helix has been wound up. FDA states that pDNA may be less effective in the open circular or linear form and thus requires almost all pDNA content in the supercoiled form for its clinical approval. However, there is not yet a consensus opinion whether the therapeutic effectiveness in gene therapy is affected by the detailed topological structures, as efficient transfections have been already described using these tertiary forms of pDNA [14].

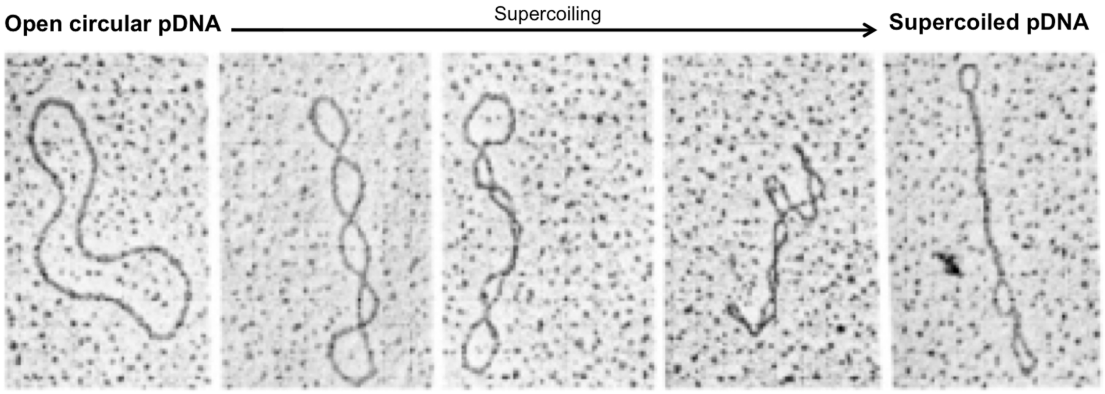


Figure 3 – Plasmid DNA supercoiling. Adapted from [52]

The open circular and linear forms result from enzymatic or shear-induced breakage of the sugar phosphate backbone and represents also a critical and difficult purification challenge since they are also treated as an impurity, but share identical nature to the supercoiled DNA [38].

Another hurdle in the pDNA downstream processing is the pDNA vector size, which may vary from 2-50 kbp dependent on the length of the therapeutic gene and/or the length of the promoters or other gene elements integrated in the construct. The extraction of large plasmids from cells without being damaged is a significant challenge for the biochemical engineers and usually does not perform well as smaller plasmids [43]. Commonly, this is a generic effect and not sequence specific. The increase viscosity and the shear sensitivity of the larger plasmids are the major concerns, but also the difficulty to separate the pDNA from other larger similar molecules such as gDNA. For these reasons, small vectors are usually preferred since they are easier to purify but also because small pDNA vectors penetrates the target cells more efficiently and also shows increased expression rates in contrast with larger vectors [53]. In addition, smaller pDNA has higher copy number, hence better product yields can be obtained [14].

In summary, from the design of plasmid to the bacterial cell fermentation followed by the pDNA extraction, there are several and distinct challenges which a biochemical engineer have to address and overcome in order to obtain

clinical grade plasmid DNA. Several approaches have been developed from the initial stage of pDNA production to the formulation/polishing step. In the next Section 1.3, the most relevant approaches to overcome the downstream processing challenges will be presented.

1.3 How to purify plasmid DNA

Several commercial kits and protocols are now available for the purification of plasmids at the laboratory scale from different manufactures (Quiagen®, Promega®, Invitrogen®, etc). However, none of them are suitable for large-scale production of plasmids due to the utilization of materials unsuitable for human use (e.g. alcohol, RNase A, chloroform) and/or are, technically difficult to scale up or economically not profitable [6]. As a result, there is an ongoing search for the design of a pDNA purification process, which address the utilization of safer materials and solvents, scalability, robustness, and efficiently removal of all contaminants/impurities.

In Figure 4 there are represented several alternative and innovative pDNA extractions strategies that have been proposed and applied from the initial stages of the pDNA downstream processing to the final polishing step. The advantages and disadvantages of each technique will be outlined in this Section, as well as principles behind each one.

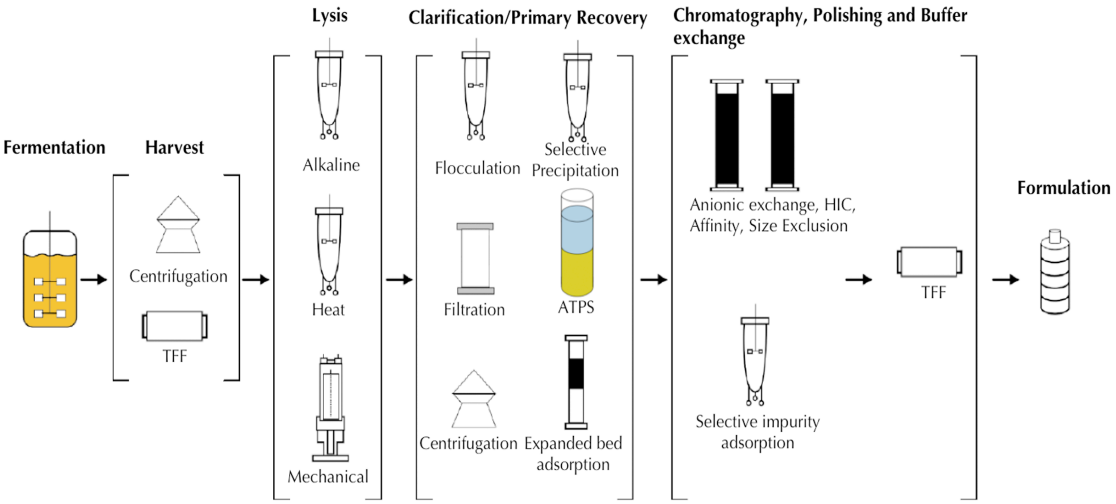


Figure 4 - Typical plasmid production alternatives. Commonly used operations for each phase of the process maybe combined to make variety of processing routes.

1.3.1 Cell recovery, lysis and removal of solids

Cell recovery processes and the cell lysis stage marks the initial steps of the downstream processing of plasmid DNA. After cell culture fermentation the pDNA-contained cells are recovered usually by centrifugation or by filtration in order to concentrate out from the growth media. Bacterial cells walls are then disrupted for the pDNA release to the supernatant in order to be captured in the following purification steps. Cell disruption fall into two main categories: mechanical and chemical disruption [16].

The mechanical disruption (e.g. sonication, bead milling, heat, etc) is usually an excluded option since the mechanical stress may potential damage the genomic DNA resulting in varied DNA fragments with the same size of the pDNA, complicating the plasmid isolation from these lysate impurities as discussed above [43].

Alternatively, a chemical lysis process has been widely adopted as a variation of the alkaline lysis originally described by Birnboim *et al.* [54]. Briefly, cells are resuspended in a neutral buffer, to which is added an alkaline 0.2 M NaOH solution containing 1 % (w/v) SDS. This step solubilises the cells walls releasing its contents to the supernatant. Within the pH range of 12.0-12.5 the pDNA does not fully denatures and remains in solution, whereas within this pH range both gDNA and proteins irreversibly denatured. This is a critical step since higher pH buffers will also irreversibly denature pDNA and may also affect the supercoiled form content of pDNA. The lysate shows marked non-newtonian properties, exhibiting a rheological behaviour that makes flow and handling of the material very difficult [43]. Such pH extremes may occur in a mechanical stirred lysis reactor, given the high concentration of NaOH and SDS. At large scale, mixing cannot be very efficient and consequently the lysate homogeneity in the vessel may be compromised and affect the pDNA quality and purification efficiency. The third step is used to lower the pH of the lysate mixture to 5.5 by addition of ice-cold sodium acetate solution. The change on the pH causes the flocculation of the gDNA and the precipitation of protein-SDS complexes and some RNA [14; 43].

The separation of the precipitated and flocculated material in the pDNA containing supernatant is usually achieved at large scale by low-shear filtration. Whereas, at the laboratory and preparative scale, centrifugation on fixed-angle rotors are the most common operations [55]. At large scale, the continuous feed flow centrifugation has been also described due to high capacity, but the liquid entering the centrifuge can cause shearing effects compromising the efficient removal of debris [56; 57]. After the removal of the debris, most gDNA and proteins are also removed, even so, the pDNA content still represents less than 3 % of the total lysate contents (see Table 1).

Some times a clarification and concentration step is performed after this first step. The clarification is mainly designed to remove remained proteins in the supernatant and usually is performed by using chaotropic salts, such as, lithium chloride and ammonium acetate [58; 59]. Plasmid DNA can be concentrated using poly(ethylene glycol) (PEG) in order to concentrate for the next purification steps but also to remove some host nucleic acids. This step has also an advantage of enabling the buffer exchange preparing the pDNA lysate for the sequent steps [14]. Alternative precipitation methods are discussed in more detail in the Section 1.3.3 – non-chromatographic purification processes.

The patent landscape in this initial lysate preparation steps includes various methods and devices to perform the alkaline lyses at large scale. A notable patent is issued to Boehringer Ingelheim, which discloses a method and device for continuous flow lysis that utilizes glass bead columns to act as lysis and clarification reactors [60]. A high salt precipitation of impurities is described in the patent WO2004060277 [61] to remove some of the RNA and much of the endotoxins.

1.3.2 Chromatography purification of pDNA

Chromatography is a well-characterized and well-established method for protein purification in the pharmaceutical industry and considered the method with the highest resolution and reproducibility [62]. It is therefore one of

techniques that has been widely explored in plasmid DNA purification systems. Several physicochemical features of either the target pDNA, or its impurities, can be exploited via selective interactions with the solid chromatographic supports. Properties such as size, charge, hydrophobicity, conformation and accessibility of molecular groups have been studied. An extensive review of the currently available chromatography matrices for the separation of plasmid DNA has been given by Diogo *et al.* [63]

1.3.2.1 Anion-exchange chromatography

Anion-exchange chromatography (AEXC) is the most used method for the primary capture of plasmid DNA with more then 81 % of the patents within DNA purification methods [64]. AEXC explores the electrostatics interactions between the negatively charged plasmid and the stationary phase bearing positively charged groups such as quaternary amines [65].

The main limitation of this technique is the co-purification of anionic polymers of similar structure and charge with pDNA, such as gDNA, endotoxins and some RNA [65; 66]. Therefore, additional purification steps are necessary, increasing the operational costs. Another disadvantages of AEX chromatography are the relative low capacity for the pDNA, and its elution in high salts buffers.

Nonetheless, some AEXC supports has been shown to be selective to the supercoiled form of pDNA (which have higher charge density), eluting later the open circular forms (which have lower charge density) [57]. This is one of the major advantages of this technique. The efficient removal of RNA (although usually with the pre-treatment of the lysate with RNase), oligonucleotides and some proteins are the others main advantages of this chromatographic method [63]. Anion-exchange stationary phases are used in most of lab kits used for the pDNA purification, such as the commercial available from Qiagen or Promega.

1.3.2.2 Size exclusion chromatography

Size exclusion chromatography (SEC) explores the different hydrodynamic size of plasmid and its impurities rather than the adsorption process in order to purify the target molecule. To achieve good separation between two species, it is required that the two molecules differ in size by 2-fold otherwise co-elution of the two species is obtained hence the resolution is limited [67]. Although good removal of impurities, such as RNA and proteins, has being observed, the separation between the pDNA and the gDNA proved to be more difficult since both nuclei acids eluted near the void volume [67], or efficient separation was dependent of the concentration of the two species in the lysate [68].

Thus, the main disadvantages of this approach are the co-elution of impurities with similar size to the pDNA; the limited scale-up to a manufacturing scale; and the high dilution factors that may result [63]. Therefore, SEC is applied more often at later stages of the pDNA purification process for the polishing step and to accommodate the pDNA on the formulation buffers.

1.3.2.3 Hydrophobic Interaction Chromatography

Hydrophobic interaction chromatography (HIC) has also been used to purify plasmid DNA [49; 69; 70]. HIC makes use of the more hydrophobic nature of single stranded nucleic acids impurities (RNA, denatured gDNA and denatured pDNA) and endotoxins to retard the flow of these impurities through a chromatography column with a hydrophobic support (*e.g.* phenyl, octyl), consequently that supercoiled DNA is eluted in the flow through [63]. HIC needs high salt concentration to promote binding, thus it can be more conveniently used after anion exchange chromatography, or when the plasmid pool has a high salt content.

Despite the low plasmid binding capacities of HIC resins, yields as high as 95 % have been reported with minimal solvent requirement [71]. The main limitation of the technique is that it requires high salt concentrations (e.g. 1.5-2.5 M ammonium sulfate) and pDNA is eluted in high salt buffer. Additionally, pDNA

has to be pre-precipitated from the lysate using isopropanol (with limited utilisation at large-scale) followed by its solubilisation in high salt concentration buffer prior to loading into the column, increasing the operation time and costs [63; 70]. Several patents have been issued using HIC for pDNA purification [72; 73; 74].

1.3.2.4 Affinity chromatography

The affinity capture of plasmid DNA based on their biological function or individual chemical structure has been developed recently and is gaining more interest in academia. Nonetheless, there is not yet an extensively literature available. At the present time there have been developed mainly three relevant approaches for the high affinity recovery of pDNA: Triple-helix, DNA binding proteins and amino acid interactions [75].

Triple-helix affinity (THAC) is based on the sequence specific interaction of a pyrimidine oligonucleotide with pDNA [76; 77]. The binding occurs via the major DNA groove and through the formation of Hoogsteen hydrogen bonds. The binding kinetics is generally slow (may need >1 h) and requires acidic pH in order to the binding complex be stable. Despite the discrimination of different plasmids based on their sequence in one step, the recovery yields are typically low (best yield ~50 % [77]) and it is necessary long chromatography runs due to the slow kinetics of triple-helix formation. However, a good elimination of the gDNA, RNA has been achieved.

The first demonstration of the usage of DNA binding proteins as affinity ligands on a pDNA chromatographic purification was performed by Woodgate *et al.* [78]. In this study the bifunctional Zinc Finger DNA-binding protein fused to glutathione S-transferase Zinc-finger (GST-ZnF) was used for the direct isolation from crude lysate of pDNA bearing the Zinc Finger recognition site. Using this approach it was possible to distinguish plasmids that contained the Zinc Finger recognition sequence from other plasmids [78]. Nonetheless, very low yield was obtained (10 %) and the authors did not demonstrate the pDNA elution from the

binding complex.

A similar strategy was developed within the same research group exploiting a different heterofunctional protein, LacI-His_6-GFP, that binds to the *lac* operon in a sequence specific manner [79; 80]. This technique allowed the isolation of a plasmid DNA bearing the *lac* operon directly from crude cell lysates free of gDNA, RNA, proteins and open circular pDNA. The major limitation of this chromatography process is the lack of capacity and difficulty to scale-up.

The two pDNA affinity proteins described above are the central affinity ligands used in aqueous two-phase systems presented in this thesis. More detailed specifications and mechanism of interaction will be further discussed in the following chapters.

Finally, amino acid–DNA affinity chromatography has been explored by using histidine [81] or arginine [82] amino acid ligands [75]. Usually referred as pseudobio affinity ligands, both these amino acids have been immobilized into a chromatography support and discriminate between supercoiled and open circular pDNA forms in one operation step. The elimination of gDNA and endotoxins were performed within the acceptable levels and no proteins or RNA was detected in the elution sample. However, these ligands have shown to have relative low yields and the interaction between pDNA and amino acid is dependent on the base composition, which may compromise future purification process with different plasmid [75].

1.3.2.4 Monoliths

Technological alternative to porous particles, monoliths, are currently gaining increase significance in the pDNA purification processes. Monoliths are single piece of macroporous (10-4000 nm or larger) solid material with high mass transfer rates, which make them more suited for larger molecules such as pDNA than traditional bead chromatography [83]. This new novel type of chromatographic stationary phase, has thus improved capacity (can reach up to

10 mg/mL) over the traditional chromatography supports (tenths of milligrams plasmid per milliliter) resulting in higher pDNA recovery yields [84].

Furthermore the optimal separation of large molecules can be achieved with short and fast runs without sacrificing resolution. Polymethacrylate based short monolithic columns distributed under the trade name "Convective Interaction Media (CIM)" (BIA Separations, patent [85]) is the most common type of monolith column being used. Several studies reporting the use of this material for pDNA purification have been published [84; 86; 87].

The difficulty in preparing high-volume monolithic columns with uniform pore structure, owing to the exothermic nature of the polymerization, is probably the major drawback of this technique. This could be circumvented by preparing individual units and merging them to form one large monolith with uniform properties, which will behave as a single radial-flow tube [62].

1.3.3 Non-chromatography processes

Alternatively, non-chromatography processes have been developed for the pDNA purification in order to circumvent some of the pDNA chromatography disadvantages. Although high resolution can generally be achieved using a chromatographic technique, the capacity for pDNA is usually low, resulting in relative low pDNA yields. This is usually attributed to the fact that chromatographic supports have been originally designed with small pore sizes for proteins purifications and not for much larger molecules such as plasmids [63; 78]. It has been observed that in these chromatographic supports the plasmid adsorb only at the beads´ outer surface and thus, not all the potential binding sites are used to bind to pDNA [51]. As a result, whereas the capacity is usually in the order of 200 mg of protein per mL of matrix, it is only about few hundreds micrograms for plasmid DNA [51].

Although new supports are being developed with high binding capacities, such as the monoliths, chromatography process are also difficult to scale-up and at

larger scales, the resolution can be lost due to diffusion problems with large volumes feeds [62]. Chromatography based process are also acknowledged to be more a expensive technique in terms of global operation costs (columns, pumps, process controller, high volume of buffers, absorbents, etc) comparatively to a non-chromatographic approach [88].

Chromatography-free processes are thus very attractive approach since it could circumvent the disadvantages of the chromatographic technique, therefore it has a high potential to improve the overall pDNA purification processes. The costs of the downstream processing of a biotech product is a major contributor to the overall process economics (estimated as high as 80 % [89]), hence an important attention is given to this stage. There is great potential for substantial savings to be made in the downstream processing by improvements of existing methods. Eliminating the need of chromatography steps will dramatically reduce these costs.

In this Section it is not intended to cover the aqueous two-phase systems (ATPSs) extraction of plasmid DNA although it is probably the must explored non-chromatographic method for the purification of therapeutic pDNA. Detail description of the technique and the current the state-of-the-art on the pDNA purification from bacterial cell lysates are given on the Section 1.4 and 1.5.

1.3.3.1 Precipitation

Precipitation techniques are a convenient and common way to separate DNA from soluble contaminants such as RNA and some proteins at the lab scale. Plasmid can be precipitated selectively by the addition of isopropanol [58], poly(ethylene glycol) [16; 61], spermidine [90], cetyltrimethylammonium bromide, CTAB [91; 92]. Curiously, CTAB was used in one of the first industrial application of a non-chromatographic process for the purification of plasmid DNA [93]. Using CTAB at 20 g/L an overall recovery of 83 % of pDNA was obtained.

Although good removal of pDNA main impurities can be achieved, the precipitation approach is not easily scaled-up since plasmid may need to be centrifuged out and the resulting pellet resuspended for further processing. Additionally, some solvents are not recommended for large-scale manufacturing since some of them are not safe for human health or may need special facilities such as spark-proof ones.

Alternatively, some precipitation methods have been developed for the selective precipitation of pDNA impurities while pDNA remains in solution. For example, RNA can be precipitated out by antichaotropic salts, such as ammonium acetate [94], ammonium sulfate [95], lithium chloride [96] but the use of high concentrations of these salts is environmentally and economically problematic at large scale.

Of notable interest are the recently developed plasmid affinity precipitation reagents such the thermoresponsive elastin-like proteins (ELPs) [97; 98]. ELPs are conjugated with a DNA-binding protein resulting in a thermally reversible affinity precipitation system, specific for plasmids that incorporate the corresponding DNA-binding sequence.

1.3.3.2 Membrane

Adsorptive membranes are another non-chromatographic technique used to reduce the concentration of plasmid impurities and an alternative option for the desalting and condition/formulation step. The main advantages of this technique are the increased loading capabilities of some of these membranes (~10 mg/mL of membrane) and the potential to maintain high efficiencies with large molecules at high flow-rates [99; 100]. Nonetheless, the relatively high cost of the membranes prevents their use at large scale [101].

Different approaches have been developed for pDNA purification. Tangential flow filtration (TFF) and nitrocellulose-integrated process have been optimized for the pDNA purification with pDNA recoveries virtually of 100 % and with

significant removal of RNA, proteins, endotoxin and gDNA [102]. Charged-based membranes have also been developed (just like in anion-exchange chromatography) for rapid and high capacity (~10 mg/mL) purification of pDNA [100; 101]. However, contaminant RNA could not be removed completely without the use of RNAse. The poor purification factor is probably the main drawback of this technique, which requires the pre-treatment of the feed stream to remove excess of RNA, either by using RNAse or extended alkaline lysis (1day) [103].

Representative examples of patents claiming the use of TFF for a pDNA purification process are issued to Butler *et al*. [104] and for Pall, for an anion exchange membrane, Pall Mustang™ [105].

1.3.4 Integration / multistage operations

Since some of the techniques described above cannot achieve high pDNA purity/quality pDNA in one single operation, usually several combinations of different techniques are explored in order to obtain the required standard product specifications. Prazeres *et al*. [65] designed several overall process flowsheets representing the current panorama in downstream processing. In all them multiple stage downstream processing was assumed for efficient overall processes.

Several examples of combined operations have been also reported elsewhere, such as the combination of AEXC, triple Helix and Hydrophobic interaction chromatography that was developed by Centelion (Gencell) yielding pDNA within the product standards [19].

Urthaler *et al*. [51] have demonstrated an automated continuous system of HIC, AEXC and SEC, delivering high pure pDNA homogeneities (98 %) of the supercoiled form in any desirable final buffer.

Cobra Bionufacturing Plc (UK) uses three chromatographic stages, starting with a anion exchange, followed by size exclusion and mixed mode

chromatography, which uses the commercial adsorbent, Poly-Flo™ that binds to endotoxins during the purification [64].

Such combinations of multi-chromatography techniques are very common, however, they are not cost effective for commercial scale purification [19]. Alternatively, the combination of chromatographic and non-chromatographic techniques in the same overall purification process is frequently developed. However, it is unusual for an overall process to eliminate the use of packed-bed chromatography totally and at least one chromatographic process is commonly necessary for the final polishing step [65].

For example, it has been performed the integration of non-chromatographic process such as the utilization of PEG precipitating agents to enhance AEXC [106]. The authors claim the use of PEG-8000 (at concentration of 1 %) prior to AEX chromatography with recoveries increased from about 20 % to 80 %.

Frequently companies have issued patents for only individual stages of the downstream processing of pDNA which have to be integrated with third-parties systems for complete downstream processing operation or for the upstream processing stage. GE Healthcare, for example, has developed the PlasmidSelect Xtra downstream purification package [107], which combines different proprietary technologies: the high fermentation process from Nature technology Corporation, and a proprietary thiophilic aromatic chromatographic media [19; 108]. S-aryl ligands represents a selective tool to remove unwanted forms of plasmid DNA, thus it is highly selective for the supercoiled pDNA. The supercoiled concentration of the affinity column reached 96 % from 45 % in the original feed [14]. Other companies, such as Boehringer Ingelheim and Merck have combinations process platforms that couple high yield fermentation, cell lysis and purification patents portfolios allowing them to perform pDNA production operation almost with own licensed technologies [19].

1.4 Aqueous Two-Phase Systems

Aqueous two-phase systems (ATPSs) are probably the most well explored non-chromatography process for the purification of plasmid DNA. In this Section it will be described in more detail the ATPS technique and the principal factors involved on the biomolecule partitioning in these systems. The current state-of-art on the utilisation of ATPS for the pDNA is also presented.

1.4.1 Advantages / Disadvantages of ATPSs

Aqueous two-phase systems are known for more than 100 years, but it was the pioneering work of the Swedish biochemist, Per-Ake Albertson started in 1956 that have demonstrated its application for biological material separation and extraction [109]. Since that time, enormous advances and new systems have been introduced and extensively studied for the recovery of cells, viruses, proteins (e.g. enzymes, antibodies), amino acids and nucleic acids [110]. Unlike proteins, that have been studied in ATPS for a long time, plasmid DNA purification using ATPS had only recently attracted more attention. Few related references have appeared in the literature before the year 2000 (only the work of Cole 1991 [111]and Ohlsson 1978 [112] has been reported), and most of the advances in this area have developed since that time.

In ATPS, the bulk of the two phases are mainly constituted by water (80–95 %), and most polymers have a stabilizing effect to the biological products. This provides a biocompatible and "friendly" environment for the bioproduct [88; 109]. In addition, the ATPS allow the integration in one step of several downstream processing steps, such as clarification, concentration and purification, with fast separations times (due to low viscosity of the phases) and technical simplicity [113; 114]. Furthermore, the scaling-up of this technique is easy and reliable and most of the traditionally equipment used in chemical industry for liquid-liquid extractions can be easily adapted for the ATPS extraction of biological products [110]. The interfacial tension is very low in

these systems, creating a high interfacial contact area of the phases, and thus it shows a good mass transfer between phases [115].

However, despite the advantages described above, ATPS are not extensively used in industrial/commercial processes. Such situation may be attributed to several factors including: the difficulty to predict the partition behaviour of a given biomolecule; the cost of the phase forming polymers; the utilization of environmental aggressive salts (e.g. phosphates); and the lack of knowledge of the technique [116]. In addition, the selectivity of ATPS is usually low and further steps are commonly needed to remove remaining product impurities. Although successful exploitation of the ATPS has been reported, the potential usage of ATPS as main purification process in the future requires that these disadvantages be overcome.

ATPSs are generally regarded as a cheaper alternative to all the chromatographic approaches for the purification of plasmid DNA [88]. However, a detailed cost comparison between currently established techniques and ATPS-based process is currently lacking. Nonetheless, as intermediate recovery unit operation, ATPS have been found to be less cost effective than other non-chromatographic techniques [117]. ATPS composed with PEG 600/ammonium sulfate (AS) has been simulated for the overall operation costs and resulted that ATPS is 2.5-fold more expensive comparatively with isopropanol/ammonium sulfate precipitation and tangential flow filtration. Nonetheless, this study, shows that pDNA yields ≥90 % were always obtained in all three techniques but ATPS delivered the highest HPLC purity (59 %), followed by precipitation method (48 %) and TFF (18 %) [117].

With the objective to lower the ATPS operations costs, cheaper and crude phase forming polymers have been considered with similar product purification efficiencies [110]. For example, efforts have also being put into replacement of the fractionated dextran (DEX) (proximally US$500/Kg [118]) by less expensive polymers, such as crude dextran, starches and maltodextrins as the bottom-phase component, which greatly reduces the processing cost [110; 119].

Alternatively, the phase forming components can be recycled, which may minimize these limitations and, at the same time, be less environmental harmful [113]. The disposal of some salts, such as phosphates, in wastewater lines are known to be toxic for the environment and its not recommended to be discarded without previously treatment [117]. The components recycling will attenuate this problems and lower the operation costs at the same time.

The poor selectivity of these systems has been addressed by using affinity ligands for the target product [120; 121]. Usually the ligand is covalently bounded to one of the phase forming polymers (usually PEG) enhancing the partitioning of the target product to the phase with less contaminants.

Several protein-based affinity ligands (Protein A [122] , antibodies [123]), or small molecule-ligands (NADH [124], ATP [125] and fatty acids [126] have been grafted to PEG and used in this way. The utilization of affinity ligands for the pDNA purification will be address in more detail in Section 1.5.

1.4.2 Phase diagrams

Aqueous two-phase systems are formed spontaneously upon mixing two aqueous solutions of structural different components above a threshold concentration [109]. In contrast to what may be expected, phase separation is a relative very common phenomenon. However, the mechanism of phase forming is not yet fully understood and sometimes apparently miscible solutions results in two-phase systems. For example, polymers very closely related and highly hydrophilic may show phase separation when mixed, such as dextran sulfate and dextran polymers [110].

It is recognized that the two principal factors to determine the result of mixing of two polymers are 1) the gain in entropy and 2) the interaction between the molecules [109; 127]. The prevalence of one these factors will determine the result of mixing: two phase separation or homogeneous solution.

A typical phase diagram is represented in Figure 5. Phase diagrams defines the potential working area for a particular ATPS and it is characteristic for a specific mixture of two polymers or polymer-salt systems under a set of conditions, such as buffer concentration, temperature, and pH. The two-phase region is defined by the binodal curve represented in Figure 5. Systems with the composition above the binodal will form two-phase system, whereas a system with the concentration of the phase components below the binodal will form a one-phase solution.

Turbidometric titration is a commonly used method for determination of the binodal since when the two phase components are mixed, the mixture become turbid only when immiscibility occurs and thus turbidometry can be measured [109]. Briefly, a series of systems of known total composition are prepared in order to make two-phase systems. Upon dilution, the mixture will eventually turn "clear" when one phase is formed. The system composition, at the point of transition, is calculated and defines the binodal curve.

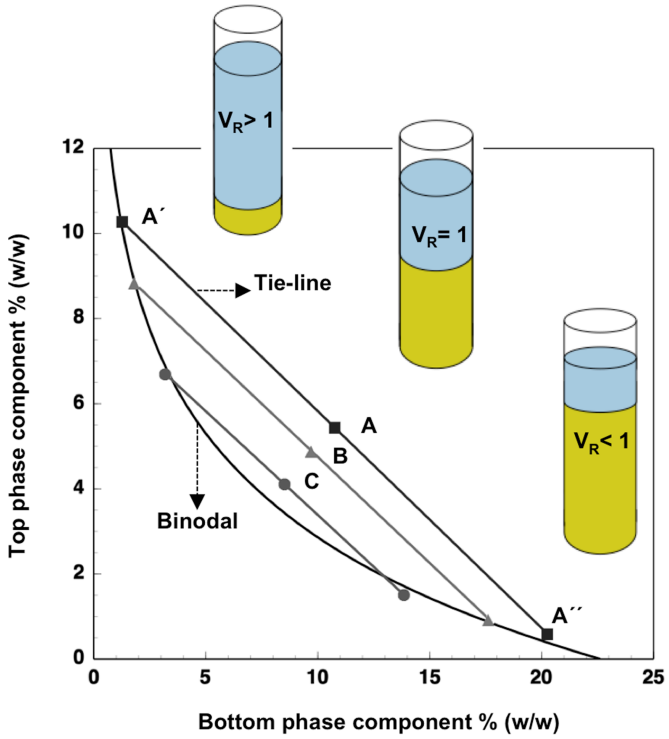


Figure 5 – Phase diagram showing the binodal curve for systems composed with two phase-forming components. Tie-lines are represented and three selected ATPS, from the same tie-line, with different volume ratio (V_R) are schematically drawn.

There are several information that can be extracted from the diagram: 1) the minimum concentration of the phase forming polymers or salts for two-phase separation; 2) concentration of phase components in the bottom and top phase; and 3) phase volumes ratio.

Suppose a point A (Figure 5) represents the total composition of a system, percentage of the top and bottom phase component per total system weight. The composition of the bottom and top phase of that particular system are the coordinates of the tie-line intersection (nodes) with the binodal curve. In Figure 5, the A´ is the composition of the two components in the top phase, whereas the A´´, is the phase composition of the bottom phase. Any other system with the total composition lying in the same tie-line, will give rise to an ATPS with the same top and bottom phase composition, but with different volumes of the two phases as schematically represented in the Figure 5. If the composition is expressed in weight per weight (w/w) percent, the weight ratio of bottom phase / top phase is equal to the ratio between lines AA´ and AA´´. Hence from the diagram:

$$\frac{V_{Top}d_{Top}}{V_{Bottom}d_{Bottom}} = \frac{\overline{AA'}}{\overline{AA''}} \qquad (1)$$

where, V is the volume of top and bottom phase; and d the densities of polymer or salt phases. Since the densities of the top and bottom are not very different (~1.0-1.1) the ratio of the segments AA´ and AA´ is proximally the volume ratio in that system.

1.4.3 Tie-Line Length

Tie-lines are determined by calculating the polymer or salt composition on both the top and bottom phases for each system studied. For example, PEG concentration (a commonly phase-forming polymer) is usually determined by refractometry using a calibration curve constructed with PEG standards of known concentration. The dextran (DEX) concentration, other traditionally used phase-forming polymer, is quantified by polarimetry using calibration curves

obtained for dextran solutions of known concentration. Salts are usually quantified by conductivity measurements [109; 128]. Tie-Line lengths (*TLL*) characterise the compositional differences between the two phases and in a system with a component 1 (C1) and a component 2 (C2) re calculated using the following formula (2):

$$TLL = \sqrt{(\Delta[C1])^2 + (\Delta[C2])^2} \tag{2}$$

where

$$\Delta[C1] = [C1]_{top} - [C1]_{bottom} \quad ; \quad \Delta[C2] = [C2]_{top} - [C2]_{bottom} \tag{3}$$

All aqueous two-phase systems compositions are expressed in % (w/w). The point where the $TLL = 0$ is called the critical point and lies on the binodal curve. This is a theoretical system where the composition and volume of both phases are equal. In theory, in this point a particle will partition equally to both phases since the composition of the two phases are similar. The critical point can be determined by drawing a line connecting the middle points of each tie-line (A, B, C in Figure 5) and extrapolated it to the binodal. The critical point is the intersection of this line with the binodal.

1.4.4 Partitioning in Aqueous Two-Phase Systems

The basis of biomolecule separation in a two-phase system is the selective distribution of substances between the two phases. It is this phenomenon of partition that is exploited in ATPS extractions operations, since it provides a method whereby the desired solute can be removed from a solution, and separated from other undesirable impurities.

The distribution of substances between each phase is governed by a number of parameters related with the properties of the phase system and the substance, and also by the interaction between the two through hydrogen bonds, van der waals, electrostatic, and hydrophobic interactions [109; 115]. The biomolecule partitioning is also influenced by conformation effects, steric hindrance and size

of the biomolecule, increasing the difficulty to predict its partitioning behavior in ATPS. Since there are many forces governing the partition, the separation modeling of a biomolecule in ATPS poses an extremely complex problem because of its dependence on the above array of factors.

The partitioning of a molecule, x, is normally defined as the partition coefficient, K, which is the ratio of the concentration of the molecule in the top phase $[x]_{top}$ to that in the bottom phase $[x]_{bottom}$:

$$Kx = \frac{[x]_{top}}{[x]_{bottom}} \tag{4}$$

The contributions of these factors to the observed partition coefficient can be summed up in logarithmic terms, as follows:

$$\ln Kx = \ln K_0 + \ln K_{elec} + \ln K_{hydrop} + \ln K_{size} \tag{5}$$

wherein $\ln K_{elec}$, denotes the electrostatic effects; $\ln K_{hydrop}$, a hydrophobic term; $\ln K_{size}$, a term for biomolecule and polymer size; while $\ln K_0$ represents other factors.

All the properties are not equally important and this depends on the system used. For example, hydrophobicity was shown to be more important in polymer-salt systems than in polymer-polymer systems [129]. On the other hand, the physicochemical properties of the phases depends on the conjugation of other factors, such type and the molecular weight of polymer, pH, salt composition, molecular weight, and temperature; adding to the complexity mechanism of partition [115].

It should be noted that the complexity of these systems is even greater because these factors are not completely independent of each other. Consequently, due to the lack of partitioning models, most of the times the partitioning of a biomolecule has to be determined experimentally and this is probably one of the reasons why this technique has not yet been widely adopted [116].

However, this also means that the partitioning of a compound is rather unique

and the modification of some parameters may increase the selectivity of that compound to the phase where fewer impurities accumulate [130]. Nonetheless, the selection of suitable ATPS could be a rather laborious process in order to find the appropriate/desirable conditions. The utilisation of robotic high throughput screening approach would be an advantage at this stage, since it is possible to test several factors in a short time, such as, system composition, pH, polymers molecular weight, ionic strength and all other added compounds [131].

Currently, there is not yet a holistic model describing the selection of key system parameters for the purification of any target biomolecule. The elucidation of the molecular mechanistic understanding of solute partitioning is currently one of the major areas of research. Despite advances that have been made, additional knowledge is needed to fully understand the partitioning mechanisms [116].

1.4.5 Aqueous Two-phase Systems for pDNA purification

The recovery of pDNA by ATPS often makes use of the polymer PEG, and a second polymer- or a salt- enriched bottom phase. Potassium phosphate, ammonium sulphate and sodium citrate had been used as phase-forming salt with comparable results [70; 114; 132]. In most of the studies published using the PEG/salt systems, high recovery of pDNA is obtained to the salt phase with the elimination of most RNA, proteins and endotoxins to the polymer phase. Better results are usually obtained using a PEG-salt system with low molecular weight PEG. However, the system composition must be carefully chosen as significant differences between the top and bottom phase compositions, corresponding to high tie-line values, result in plasmid precipitation at the interface, with the concomitant yield reduction [70].

The PEG molecular weight in PEG/$(NH_4)_2SO_4$ systems seems to have greater effect on the pDNA partition. Results showed that pDNA could be directed towards the top phase (PEG-rich phase) (MW<400) or to the bottom phase (salt-rich phase) (MW>400) [70; 132]. In PEG 600/salt systems the pDNA partitions

totally to the bottom phase (yield 88.4 %) whereas the protein from bacterial cell lysate remained in the top phase [70].

Rahimpour *et al*. [133] reported 99 % pDNA recovery in systems composed with PEG400/citrate at pH 6.9. Under these conditions the RNA removal was only about 68 %. A recent study reported that while ammonium sulphate systems give higher purity (~100 %) compared to the citrate systems, a lower recovery yield is obtained [134]. A compromise between these two parameters could be found using a mixture of the two salts. A mixture of 25 % (w/w) ammonium sulfate and 75 % (w/w) sodium citrate resulted in plasmid recovery of 91.1 % and purity (17.2 %). In this study it was demonstrated that ammonium sulfate could be partially replace with citrate buffers without compromising performance [134]. However, purification improvements are still needed in order to increase the pDNA purity.

From an environmental point of view, the utilisation of citrate is preferable to sulfate or phosphate systems, because its biodegradable and non-toxic [117; 133]. On the other hand, the extraction of pDNA to the ammonium sulfate phase is advantageous since the salt phase could be applied directly to HIC chromatography for removal of remained impurities [70].

Polymer/polymer systems are usually regarded as a more expensive alternative than polymer/salts systems [88]. However these systems can avoid some of problems associated with the polymer/salt systems, such as the interference of the salts on the utilization of ligands for the affinity partitioning of plasmid DNA [128]. In addition, there is less variation of the pDNA partition and its impurities from system-to-system in polymer-polymer systems as compared to the polymer-salts as demonstrated in this thesis.

Thermoseparating phase components such as the 50 % ethylene oxide – 50 % propylene oxide copolymer (EOPO) have been used with dextran to form two-phase polymer systems [135]. In EOPO-dextran systems almost 100 % of pDNA accumulated in the EOPO rich phase, which can be removed and heated to 55 ºC to form a second aqueous phase to eluted the pDNA. With this

approach the removal of RNA is not so efficient, although 80 % reduction of this contaminant has been achieved [135]. Another drawback of this approach is that the lysate has to be previously desalted in order to precisely control the salt composition of the systems, which determines the partitioning of nucleic acids. In contrast, in salt-polymer systems the phase components could be directly added to the lysate [114].

Some of the aforementioned studies applied other purification steps to recover the pDNA from the salt phase of ATPS and to remove remained impurities. ATPS integration with HIC chromatography [70], membranes [136] or with more than two unit operations have been reported for this purpose [137].

Kepka *et al.* [137] integrated aqueous two-phase systems with membrane filtration and lid bead chromatography, a procedure that has also been described in patent WO2004020629 [138]. The process started with a volume reduction and lysate buffer exchanged using hollow fibber membrane system. The concentrated samples were then applied to EOPO-dextran system to remove most of the proteins and RNA. Finally, the eluted pDNA from the top phase was polish using lid bead chromatography. This new type of chromatography uses positively charged inner cores (for the binding of remained RNA) and inert surface layer. A good removal of all impurities was accomplished but the overall process pDNA yield was about 69 %.

Frerix *el al.* [139] have used membranes as polishing step after a PEG-Salt ATPS extraction (avoiding totally a chromatographic step) to select the supercoiled form of pDNA (35 mg / 95 %) free from remaining nucleic acid impurities, such as open circular pDNA (<3 %) and gDNA (<1 %). This process has been also described in a patent issued in 2006, US2006286080 [140].

Another type of process integration is the combination of downstream separation operations into single separation step by adding an affinity component to the system in order to increase selectivity of ATPS. In fact, this was our approach for the development of efficient pDNA purification process.

1.5 Affinity purification of plasmid DNA in ATPS

1.5.1 Affinity-based separations in ATPS

The results obtained so far using ATPS for the pDNA purification are promising since most of impurities could be removed in a single step. However, the ATPS methods hitherto have been used mainly as primary recovery/enrichment step after cell lysis and further purifications steps are required in order to obtain clinical grade pDNA. The main limitation of using ATPS as main downstream operation is due to its low selectivity for the target pDNA. The utilisation of affinity ligands on aqueous two-phase systems extractions is thus an attractive approach since it will improve the poor selectivity of these systems and, at the same time, it will exploit the technical advantages of using ATPS for large-scale operations (see Section 1.4.1).

The affinity approach is based on the specificity and biorecognition properties of the ligands for the target pDNA. Affinity ligands has the potential for the efficient capture of pDNA from a crude feedstock with large and structurally related impurities, such as, high molecular weight RNA, endotoxin and shear genomic DNA [75].

Affinity-based separations have been extensively applied for proteins extractions in chromatography [141; 142; 143] and in aqueous two-phase systems [144; 145; 146], but only one recent study from our group, explored the utilisation of a non-specific ligand in aqueous two-phase systems for the plasmids extractions from a bacterial cell lysates [128].

Duarte *et al.* [128] have used the cationic polymer poly(ethyleimine) (PEI) as non-specific ligand for the purification of pDNA from crude cell lysates in a two-step ATPS process. The authors report to 100 % pDNA recoveries in a form of DNA/PEI polyplexes with only minor protein impurities. An initial system with the composition of PEG 600 – ammonium sulfate was used to remove most of the impurities to the top phase. The bottom phase was collected and mixed in a second system composed by PEG 3350 – DEX 110 and 0.2 % of PEI. The

purification of pDNA as polyplex form would be an advantage since after some polishing steps the polyplex would be ready for human administration. Nonetheless, safety concerns may rise in the future using this purification approach since it may be difficult to control the ratio pDNA/PEI in each operation resulting in product heterogeneity. In this study, the authors did not demonstrate the DNA elution from the polyplexes and it is not clear yet if the pDNA could be eluted and used as non-complex form. Additionally, the PEI/DNA interactions are mainly electrostatic, thus there is a potential risk for non-specificity binding of the cationic polymer to other negative charged molecules, such as RNA and gDNA, which will present some safety concerns. However, it is notable that in this study that no RNA was co-purified in system composed with 0.2 - 0.5M ammonium sulfate [128].

The utilisation of pDNA sequence specific ligands is consequently a logical approach, but yet they have not been explored in the two-phase systems so far. In this work, two protein-based ligands were used as basis for the designing a pDNA affinity purification process (Figure 6).

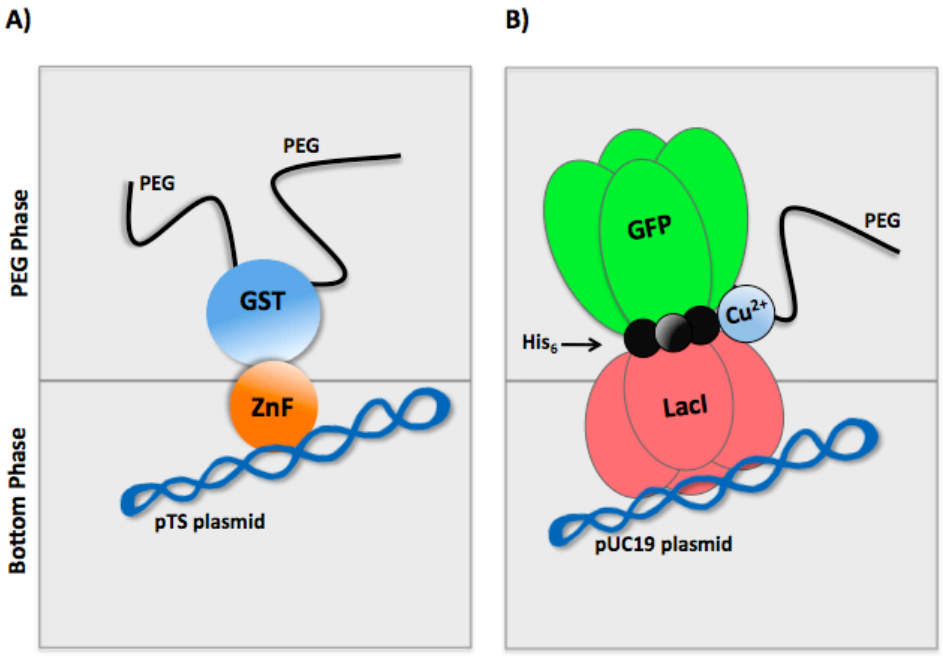


Figure 6 – Schematic representation of the two protein-based affinity ligands used in this work. A) GST-ZnF fusion protein; B) LacI-His_6-GFP

One of the ligands used is a Zinc Finger protein (ZnF) fused to the glutathione-S-transferanse (GST) protein resulting in the fusion protein named GST-ZnF [78]; the other protein ligand is a *lac* repressor protein, LacI, fused to both a His_6 tag and to a green fluorescence protein (GFP) tag resulting in the heterofunction protein named LacI-His_6-GFP [79]. These two proteins have already been proved to be efficient ligands when immobilized on to a chromatographic support for pDNA purification (Section 1.3.2.4). However, the pDNA capacity of the chromatographic process was shown to be suboptimal for large-scale pDNA production [78; 79].

The utilisation of these ligands in ATPS should thus increase the previously found low yield as the interaction takes place in an homogeneous phase and there is no steric hindrance caused by the binding to the resin. Moreover, loss of products due to non-specific adsorption to the solid support should also be avoided.

The main drawbacks of protein-based ligands are related to the fact that ligands come from a biological source and they may require previous purification and can be costly to produce. In addition they maybe contaminated with other impurities, which could complicate the downstream processing of pDNA. The utilisation of the crude cell lysates of these ligands, instead of using the pure protein ligands, could significantly reduce the high costs associated to the ligands. Nonetheless, this could increase the purification challenge since impurities will also increase in the two-phase systems.

The stability of the protein ligands is another issue of concern, as proteins are usually not very stable at room temperature for long times. Even so, it has been already shown that these ligands could be lyophilized, stored at room temperature and re-hydrated when required [80]. Alternatively, synthetic ligands can potentially be a more suitable alternative in the future, as they are more stable than the biological counterparts, easier to scale-up and they are more durable [147]. Synthetic ligands may prove to be more cost-effective, but they may lack the bioselectivity of the sequence-specific DNA protein ligands.

1.5.2 Systems selection for affinity partition

The successful and efficient utilisation of affinity ligands in ATPSs requires the selection of a two-phase system where the ligand/pDNA complex can partition into the phase with less impurities. Thus, a suitable ATPS would be one that all pDNA impurities accumulates mainly in one phase in order to be possible to extract the pDNA to the "clean" phase, therefore increasing pDNA purity.

It has been reported that in some polymer-polymer systems, particularly PEG-dextran, it is easier to select conditions where the majority of pDNA impurities can accumulate in one particular phase. ATPSs have been described in which 97.5 % of denatured genomic DNA [112]; the majority of the total proteins [109], nearly 100 % RNA and ~100 % pDNA from bacterial cell lysate accumulate in the bottom dextran rich phase [128]. We thus hypothesised that the PEG-dextran would be potential suitable systems for first extractions systems since pDNA could be extracted from its main impurities by means of affinity ligands, which would accumulate in the opposite top phase (PEG-rich phase).

However, the further utilisation of affinity ligands requires a complete understanding of the partition of the bacterial cell lysate's components on this polymer-polymer systems and the evaluation of the influence of parameters such as polymers molecular weight and concentration might have on the partitioning of these components. There are few examples reporting to the partition of DNA bacterial cell lysate in PEG-dextran systems in the literature [112; 128], and a systematic study of the partitioning of plasmid DNA from a bacterial lysate in PEG-dextran had not been made yet.

For the utilisation of affinity ligands in ATPS the polymer-polymer systems are preferred over polymer–salt systems since a high salt concentration may impair the interaction between the protein ligands and pDNA and thus should be avoid as first operation systems [128].

1.5.3 Affinity ligands relevant for this work

1.5.3.1 Zinc finger protein

Zinc finger proteins are one of the best-known class of DNA-binding proteins that binds sequence specifically to the DNA recognition site [148; 149]. A recent model proposes that these proteins bind initially non-specifically by electrostatic interaction and then run along the DNA strand until they find the recognition site where they strongly attached [150]. In this work we made use of ZnF protein which the sequence specificity has been previously described by Desjarlais *et al.* [151]. The authors have determined that the ZnF bind to the target sequence (5´-GGGGCGGCT-3´) with a K_D of 2nM and the affinity could drop substantially (to K_D of 1μM) if two bases were modified on the target sequence. Woodgate *et al.* [78] have also demonstrated that the zinc finger protein (10.7 kDa) fused to a glutathione-*S*-transferase, GST-ZnF (38.3 kDa), could distinguish plasmid that contained the target DNA sequences (pTS plasmid) from other plasmids, pUC19, that shares natively 7 bp of the 9bp. In the above study, the pTS plasmid was constructed by inserting a synthetic oligonucleotide cassette containing the recognition sequence into the *sma*I site of pUC19 plasmid.

The same bifunctional affinity protein, GST-ZnF (hydrodynamic radii of less than 10 nm) will be accessed here for pDNA extraction in ATPS systems. The GST moiety will not be removed from the GST-ZnF protein since it has been shown that the GST does not interfere with the ZnF binding properties [78; 152]. The GST-tag protein is used for the purification of the fusion protein by selectively absorption onto a glutathione affinity matrix allowing its purification from the expression host.

The main problem associate with the GST-ZnF fusion protein is related to the strong binding to the target DNA, consequently the elution from the protein-DNA complex was impractical [78]. Consequently, this ligand cannot effectively be used in an efficient pDNA purification process since the presence of the bounded protein would be restricted by the regulatory agencies. Hence, in this thesis the GST-ZnF fusion protein is used to demonstrate the proof-of-

concept of a specific affinity strategy for pDNA purification in aqueous two-phase systems.

1.5.3.2 Lac repressor protein

The repression of *lac* operon is achieved by halting the binding of the RNA polymerase in a pre-transcriptional complex with the promoter by the first lac operator (*lac*O_1) being bound to the *lac* repressor [153]. Native LacI protein is a tetrameric protein (two dimer units) that binds to the first *lac* operator, *lac*O_1, and with lower affinity to the *lac*O_2, or *lac*O_3 operators sequences [154; 155]. The binding to a second operator was found to make no significant contribution to the repression, however, stabilizes the protein/DNA complex and aids to the formation of the repression loop [79]. However, unsurprisingly, the interaction between the plasmid containing the *lac*O_1 and *lac*O_3, such as pUC19, is stronger when two rather than one operator, *lac*O_1, is present (K_D of 12 pM and 1nM, respectively) [156].

The LacI-His_6-GFP protein has the molecular weight of 274 kDa [79] and 3 domains. The GFP domain enables the specific detection of the fusion protein by fluorescence in a mixture of other proteins; the His_6 tag enables the binding to the immobilized metal affinity matrix; and the LacI domain binds sequence specifically to the *Lac*O_3/ *Lac*O_1 operator on the "native" pUC19 plasmids (~1.8 MDa).

Recently, LacI-His_6-GFP DNA binding protein was immobilized into an immobilized metal-ion chromatography column and explored for the affinity capture of pDNA directly from crude cell lysates (see Section 1.3.2.4). The resulted pDNA eluate was virtually free from all main pDNA impurities in highly pure pDNA free from contamination of gDNA, denatured pDNA, RNA and proteins.

1.5.4 Ligand PEGylation

It was hypothesised that partitioning of the DNA ligand to the PEG-rich phase could be accomplished by either (1) tailoring the two-phase PEG-dextran system such that the affinity protein preferentially accumulates in the PEG phase, or (2)

PEGylating the ligand to cause preferential partitioning into the PEG phase. The latter approach was investigated by two different methods. The first method was investigated using the GST-ZnF protein where the PEGylation was achieved by covalent binding of the PEG polymer into the affinity protein by reacting it with different activated PEGs targeted for either amine or cysteins groups within the protein. The second PEGylation method hypothesised was composed of immobilized metal–ion affinity PEG which binds to the poly(histidines) that are accessible on the surface of LacI-His_6-GFP protein.

In both methods it is expected enhancement of partitioning of the affinity protein / pDNA complex to the PEG-rich top phase, where less pDNA impurities accumulate. The DNA binding affinity ligands are dependent on the presence of recognition sequence (less then 120 bp), so the success of the systems does not depends on the size or charge of the DNA (as in PEI systems). This would enable the specifically capture of pDNA in one step without gDNA, RNA and proteins, thus eliminating the need of further purification or polishing steps.

1.6 References

[1]J.D. Watson, and F.H. Crick, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171 (1953) 737-8.

[2]D.A. Jackson, R.H. Symons, and P. Berg, Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. Proc Natl Acad Sci U S A 69 (1972) 2904-9.

[3]S.N. Cohen, and A.C. Chang, Recircularization and autonomous replication of a sheared R-factor DNA segment in Escherichia coli transformants. Proc Natl Acad Sci U S A 70 (1973) 1293-7.

[4] International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. Nature 431 (2004) 931-45.

[5]J.P. Gillet, B. Macadangdang, R.L. Fathke, M.M. Gottesman, and C. Kimchi-Sarfaty, The development of gene therapy: from monogenic recessive disorders to complex diseases such as cancer. Methods Mol Biol 542 (2009) 5-54.

[6]G.N. Ferreira, G.A. Monteiro, D.M. Prazeres, and J.M. Cabral, Downstream processing of plasmid DNA for gene therapy and DNA vaccine applications. Trends Biotechnol 18 (2000) 380-8.

[7]M.A. Kutzler, and D.B. Weiner, DNA vaccines: ready for prime time? Nat Rev Genet 9 (2008) 776-88.

[8]G.J. Prud'homme, DNA vaccination against tumors. J Gene Med 7 (2005) 3-17.

[9]A. El-Aneed, An overview of current delivery systems in cancer gene therapy. J Control Release 94 (2004) 1-14.

[10]Y. Yi, S.H. Hahm, and K.H. Lee, Retroviral gene therapy: safety issues and possible solutions. Curr Gene Ther 5 (2005) 25-35.

[11]S.E. Parker, H.L. Vahlsing, L.M. Serfilippi, C.L. Franklin, S.G. Doh, S.H. Gromkowski, D. Lew, M. Manthorpe, and J. Norman, Cancer gene therapy using plasmid DNA: safety evaluation in rodents and non-human primates. Hum Gene Ther 6 (1995) 575-90.

[12]D. Luo, and W.M. Saltzman, Synthetic DNA delivery systems. Nat Biotechnol 18 (2000) 33-7.

[13]Wiley, Gene Therapy Clinical Trials Worldwide http://www.wiley.co.uk/genetherapy/clinical/. (2009).

[14]P.A. Shamlou, Scaleable processes for the manufacture of therapeutic quantities of plasmid DNA. Biotechnol Appl Biochem 37 (2003) 207-18.

[15]W.W. Nichols, B.J. Ledwith, S.V. Manam, and P.J. Troilo, Potential DNA vaccine integration into host cell genome. Ann N Y Acad Sci 772 (1995) 30-9.

[16]D.M. Prazeres, G.N. Ferreira, G.A. Monteiro, C.L. Cooney, and J.M. Cabral, Large-scale production of pharmaceutical-grade plasmid DNA for gene therapy: problems and bottlenecks. Trends Biotechnol 17 (1999) 169-74.

[17]J.A. Wolff, R.W. Malone, P. Williams, W. Chong, G. Acsadi, A. Jani, and P.L. Felgner, Direct gene transfer into mouse muscle in vivo. Science 247 (1990) 1465-8.

[18]H. Ying, L. Shan, H. Jenny, K. Michael, and M.F. Gareth, Using DNA as a drug-Bioprocessing and delivery strategies Chem Eng Res Des Volume 87 (2009) 343-348

[19]A.E. Carnes, and J.A. Williams, Plasmid DNA manufacturing technology. Recent Pat Biotechnol 1 (2007) 151-66.

[20]R.G. Crystal, Transfer of genes to humans: early lessons and obstacles to success. Science 270 (1995) 404-10.

[21]G.M. Forde, Rapid-response vaccines--does DNA offer a solution? Nat Biotechnol 23 (2005) 1059-62.

[22]S. Manoj, L.A. Babiuk, and S. van Drunen Littel-van den Hurk, Approaches to enhance the efficacy of DNA vaccines. Crit Rev Clin Lab Sci 41 (2004) 1-39.

[23]M.P. Kieny, A. Costa, J. Hombach, P. Carrasco, Y. Pervikov, D. Salisbury, M. Greco, I. Gust, M. LaForce, C. Franco-Paredes, J.I. Santos, E. D'Hondt, G. Rimmelzwaan, R. Karron, and K. Fukuda, A global pandemic influenza vaccine action plan. Vaccine 24 (2006) 6367-70.

[24]U. Nations, Population division, 2008 revision of World Population Prospects, 2008. (2008).

[25]Ministério da Saúde http://www.dgs.pt/. (2009).

[26]Government of Canada http://canada.gc.ca/home.html. (2009).

[27]United Kingdom Goverment - Department of Health http://www.dh.gov.uk/en/index.htm, 2009.

[28]Ministry of Health Israel http://www.health.gov.il/english/. (2009).

[29]J.K. Räty, J.T. Pikkarainen, T. Wirth, and S. Ylä-Herttuala, Gene Therapy: The First Approved Gene-Based Medicines, Molecular Mechanisms and Clinical Indications. Current Molecular Pharmacology 1 (2008) 13-23.

[30]J.A. Williams, A.E. Carnes, and C.P. Hodgson, Plasmid DNA vaccine vector design: impact on efficacy, safety and upstream production. Biotechnol Adv 27 (2009) 353-70.

[31]USPTO, United States Patent and Trademark Office: http://www.uspto.gov/, 2009.

[32]A. Tejeda-Mansir, and R.M. Montesinos, Upstream processing of plasmid DNA for vaccine and gene therapy applications. Recent Pat Biotechnol 2 (2008) 156-72.

[33]K.J. Prather, S. Sagar, J. Murphy, and M. Chartrain, Industrial scale production of plasmid DNA for vaccine and gene therapy: plasmid design, production, and purification. Enzyme and Microbial Technology 33 (2003) 865-883.

[34]S.Y. Lee, K.S. Yim, H.N. Chang, and Y.K. Chang, Construction of plasmids, estimation of plasmid stability, and use of stable plasmids for the production of poly(3-hydroxybutyric acid) by recombinant Escherichia coli. J Biotechnol 32 (1994) 203-11.

[35]L. Vidal, J. Pinsach, G. Striedner, G. Caminal, and P. Ferrer, Development of an antibiotic-free plasmid selection system based on glycine auxotrophy for recombinant protein overproduction in Escherichia coli. J Biotechnol 134 (2008) 127-36.

[36]P. Hagg, J.W. de Pohl, F. Abdulkarim, and L.A. Isaksson, A host/plasmid system that is not dependent on antibiotics and antibiotic resistance genes for stable plasmid maintenance in Escherichia coli. J Biotechnol 111 (2004) 17-30.

[37]J. Luke, A.E. Carnes, C.P. Hodgson, and J.A. Williams, Improved antibiotic-free DNA vaccine vectors utilizing a novel RNA based plasmid selection system. Vaccine (2009).

[38]W.J. Kelly, Perspectives on plasmid-based gene therapy: challenges for the product and the process. Biotechnol Appl Biochem 37 (2003) 219-23.

[39]R.D. O'Kennedy, C. Baldwin, and E. Keshavarz-Moore, Effects of growth medium selection on plasmid DNA production and initial processing steps. J Biotechnol 76 (2000) 175-83.

[40]M. Chartrain, L.K. Bentley, B.A. Krulewicz, K.M. Listner, and W. Sun, Lee, C.B.: WO05078115, 2005.

[41]H. Huber, G. Weigl, and W. Buchinger, Patent ID: WO2005097990 (2005).

[42]J. Stadler, R. Lemmens, and T. Nyhammar, Plasmid DNA purification. J Gene Med 6 Suppl 1 (2004) S54-66.

[43]M.S. Levy, R.D. O'Kennedy, P. Ayazi-Shamlou, and P. Dunnill, Biochemical engineering approaches to the challenges of producing pure plasmid DNA. Trends Biotechnol 18 (2000) 296-305.

[44]R. Karima, S. Matsumoto, H. Higashi, and K. Matsushima, The molecular pathogenesis of endotoxic shock and organ failure. Mol Med Today 5 (1999) 123-32.

[45]C. Hirayama, and M. Sakata, Chromatographic removal of endotoxin from protein solutions by polymer particles. J Chromatogr B Analyt Technol Biomed Life Sci 781 (2002) 419-32.

[46]S. Liu, R. Tobias, S. McClure, G. Styba, Q. Shi, and G. Jackowski, Removal of endotoxin from recombinant protein preparations. Clin Biochem 30 (1997) 455-63.

[47]F.B. Anspach, Endotoxin removal by affinity sorbents. J Biochem Biophys Methods 49 (2001) 665-81.

[48]R. Lemmens, U. Olsson, T. Nyhammar, and J. Stadler, Supercoiled plasmid DNA: selective purification by thiophilic/aromatic adsorption. J Chromatogr B Analyt Technol Biomed Life Sci 784 (2003) 291-300.

[49]M.M. Diogo, J.A. Queiroz, G.A. Monteiro, S.A. Martins, G.N. Ferreira, and D.M. Prazeres, Purification of a cystic fibrosis plasmid vector for gene therapy using hydrophobic interaction chromatography. Biotechnol Bioeng 68 (2000) 576-83.

[50]J. Briggs, and P.R. Panfili, Quantitation of DNA and protein impurities in biopharmaceuticals. Anal Chem 63 (1991) 850-9.

[51]J. Urthaler, W. Buchinger, and R. Necina, Improved downstream process for the production of plasmid DNA for gene therapy. Acta Biochim Pol 52 (2005) 703-11.

[52]G.N.M. Ferreira, Chromatographic Approaches in the Purification of Plasmid DNA for Therapy and Vaccination. Chemical Engineering & Technology 28 (2005) 1285-1294.

[53]C. Bloquel, E. Fabre, M.F. Bureau, and D. Scherman, Plasmid DNA electrotransfer for intracellular and secreted proteins expression: new methodological developments and applications. J Gene Med 6 Suppl 1 (2004) S11-23.

[54]H.C. Birnboim, and J. Doly, A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7 (1979) 1513-23.

[55]I. Theodossiou, I.J. Collins, J.M. Ward, O.R.T. Thomas, and P. Dunnill, The processing of a plasmid-based gene from E. Coli. Primary recovery by filtration. Bioprocess engineering 16 (1997) 175-183

[56]G.N. Ferreira, J.M. Cabral, and D.M. Prazeres, Development of process flow sheets for the purification of supercoiled plasmids for gene therapy applications. Biotechnol Prog 15 (1999) 725-31.

[57]D.M. Prazeres, T. Schluep, and C. Cooney, Preparative purification of supercoiled plasmid DNA using anion-exchange chromatography. J Chromatogr A 806 (1998) 31-45.

[58]A. Chakrabarti, S. Sitaric, and S. Ohi, A procedure for large-scale plasmid isolation without using ultracentrifugation. Biotechnol Appl Biochem 16 (1992) 211-5.

[59]N.A. Horn, J.A. Meek, G. Budahazi, and M. Marquet, Cancer gene therapy using plasmid DNA: purification of DNA for human clinical trials. Hum Gene Ther 6 (1995) 565-73.

[60]J. Urthaler, R. Necina, C. Ascher, and H. Woehrer, Patent ID: WO04085643. (2004).

[61]G. Budahazi, and B. Goff, Patent ID: WO04060277 (2004).

[62]T.M. Przybycien, N.S. Pujar, and L.M. Steele, Alternative bioseparation operations: life beyond packed-bed chromatography. Curr Opin Biotechnol 15 (2004) 469-78.

[63]M.M. Diogo, J.A. Queiroz, and D.M. Prazeres, Chromatography of plasmid DNA. J Chromatogr A 1069 (2005) 3-22.

[64]G.M. Forde, Plasmid DNA purification by affinity methods. Doctor of philosophy dissertation (2004) University of Cambridge (United Kingdom).

[65]D.M.F. Prazeres, and G.N. Ferreira, Design of flowsheets for the recovery and purification of plasmids for gene therapy and DNA vaccination Chemical Engineering and Processing 43 (2004) 609-624.

[66]I.P. Wicks, Bacterial lipopolysaccharides copurifies with plasmid DNA: Implications for animal models and human gene therapy. Hum. Gene Ther. 6 (1995) 317–323.

[67]L.Z. Li, Y. Liu, M.S. Sun, and Y.M. Shao, Effect of salt on purification of plasmid DNA using size-exclusion chromatography. J Chromatogr A 1139 (2007) 228-35.

[68]M. Bywater, R. Bywater, and L. Hellman, A novel chromatographic procedure for purification of bacterial plasmids. Anal Biochem 132 (1983) 219-24.

[69]M.M. Diogo, J.A. Queiroz, G.A. Monteiro, and D.M. Prazeres, Separation and analysis of plasmid denatured forms using hydrophobic interaction chromatography. Anal Biochem 275 (1999) 122-4.

[70]I.P. Trindade, M.M. Diogo, D.M. Prazeres, and J.C. Marcos, Purification of plasmid DNA vectors by aqueous two-phase extraction and hydrophobic interaction chromatography. J Chromatogr A 1082 (2005) 176-84.

[71]M.M. Diogo, S.C. Ribeiro, J.A. Queiroz, G.A. Monteiro, P. Perrin, N. Tordo, and D.M. Prazeres, Scale-up of hydrophobic interaction chromatography for the purification of a DNA vaccine against rabies. Biotechnol. Lett. 22 (2000) 1397-1400.

[72]N. Ramasubramanyan, Patent ID: US20056953686 (2005).

[73]N. Ramasubramanyan, Patent ID: WO0073318 (2000).

[74]D.F.M. Prazeres, F. Rodrigues, M.M. Diogo, and Queiroz, Patent ID: US20077169917 (2007).

[75]F. Sousa, D.M. Prazeres, and J.A. Queiroz, Affinity chromatography approaches to overcome the challenges of purifying plasmid DNA. Trends Biotechnol 26 (2008) 518-25.

[76]P. Wils, V. Escriou, A. Warnery, F. Lacroix, D. Lagneaux, M. Ollivier, J. Crouzet, J.F. Mayaux, and D. Scherman, Efficient purification of plasmid DNA for gene transfer using triple-helix affinity chromatography. Gene Ther 4 (1997) 323-30.

[77]T. Schluep, and C.L. Cooney, Purification of plasmids by triplex affinity interaction. Nucleic Acids Res 26 (1998) 4524-8.

[78]J. Woodgate, D. Palfrey, D.A. Nagel, A.V. Hine, and N.K. Slater, Protein-mediated isolation of plasmid DNA by a zinc finger-glutathione S-transferase affinity linker. Biotechnol Bioeng 79 (2002) 450-6.

[79]R.A. Darby, and A.V. Hine, LacI-mediated sequence-specific affinity purification of plasmid DNA for therapeutic applications. FASEB J 19 (2005) 801-3.

[80]R.A. Darby, G.M. Forde, N.K. Slater, and A.V. Hine, Affinity purification of plasmid DNA directly from crude bacterial cell lysates. Biotechnol Bioeng 98 (2007) 1103-8.

[81]F. Sousa, S. Freitas, A.R. Azzoni, D.M. Prazeres, and J. Queiroz, Selective purification of supercoiled plasmid DNA from clarified cell lysates with a single histidine-agarose chromatography step. Biotechnol Appl Biochem 45 (2006) 131-40.

[82]F. Sousa, D.M. Prazeres, and J.A. Queiroz, Binding and elution strategy for improved performance of arginine affinity chromatography in supercoiled plasmid DNA purification. Biomed Chromatogr 23 (2009) 160-5.

[83]A. Strancar, A. Podgornik, M. Barut, R. Necina, T. Scheper, and R. Freitag, Advances in Biochemical Engineering/Biotechnology, Modern Advances in Chromatography, Springer–Verlag, Heidelberg, 2002.

[84]K. Branovic, D. Forcic, J. Ivancic, A. Strancar, M. Barut, T. Kosutic Gulija, R. Zgorelec, and R. Mazuran, Application of short monolithic columns for fast purification of plasmid DNA. J Chromatogr B Analyt Technol Biomed Life Sci 801 (2004) 331-7.

[85]R. Necina, J. Urthaler, A. Strancar, J. Jancar, M. Merhar, M. Barut, and A. Podgornik, Patent ID: WO03051483 (2003).

[86]J. Urthaler, R. Schlegl, A. Podgornik, A. Strancar, A. Jungbauer, and R. Necina, Application of monoliths for plasmid DNA purification development and transfer to production. J Chromatogr A 1065 (2005) 93-106.

[87]A. Hanora, I. Savina, F.M. Plieva, V.A. Izumrudov, B. Mattiasson, and I.Y. Galaev, Direct capture of plasmid DNA from non-clarified bacterial lysate using polycation-grafted monoliths. J Biotechnol 123 (2006) 343-55.

[88]A.M. Azevedo, P.A. Rosa, I.F. Ferreira, and M.R. Aires-Barros, Chromatography-free recovery of biopharmaceuticals through aqueous two-phase processing. Trends Biotechnol 27 (2009) 240-7.

[89]A.C. Roque, C.R. Lowe, and M.A. Taipa, Antibodies and genetically engineered related molecules: production and purification. Biotechnol Prog 20 (2004) 639-54.

[90]D.V. Mourich, M.W. Munks, J.C. Murphy, R.C. Willson, and A.B. Hill, Spermine compaction is an efficient and economical method of producing vaccination-grade DNA. J Immunol Methods 274 (2003) 257-64.

[91]M. Ishaq, B. Wolf, and C. Ritter, Large-scale isolation of plasmid DNA using cetyltrimethylammonium bromide. Biotechniques 9 (1990) 19-20, 22, 24.

[92]R.J. Lander, M.A. Winters, F.J. Meacle, B.C. Buckland, and A.L. Lee, Fractional precipitation of plasmid DNA from lysate by CTAB. Biotechnol Bioeng 79 (2002) 776-84.

[93]R.J. Lander, M.A. Winters, and F.J. Meacle, Patent ID: US20046797476 (2004).

[94]J.K. Evans, P. Troilo, and B.J. Ledwith, Simultaneous purification of RNA and DNA from liver using sodium acetate precipitation. Biotechniques 24 (1998) 416-8.

[95]R.B. Sparks, and J.H. Elder, A simple and rapid procedure for the purification of plasmid DNA using reverse-phase C18 silica beads. Anal Biochem 135 (1983) 345-8.

[96]G. Cathala, J.F. Savouret, B. Mendez, B.L. West, M. Karin, J.A. Martial, and J.D. Baxter, A method for isolation of intact, translationally active ribonucleic acid. DNA 2 (1983) 329-35.

[97]U.L. Lao, J. Kostal, A. Mulchandani, and W. Chen, Affinity purification of plasmid DNA by temperature-triggered precipitation. Nat Protoc 2 (2007) 1263-8.

[98]J. Kostal, A. Mulchandani, and W. Chen, Affinity purification of plasmid DNA by temperature-triggered precipitation. Biotechnol Bioeng 85 (2004) 293-7.

[99]M.S. Levy, I.J. Collins, J.T. Tsai, P.A. Shamlou, J.M. Ward, and P. Dunnill, Removal of contaminant nucleic acids by nitrocellulose filtration during pharmaceutical-grade plasmid DNA processing. J Biotechnol 76 (2000) 197-205.

[100]M.A. Teeters, S.E. Conrardy, B.L. Thomas, T.W. Root, and E.N. Lightfoot, Adsorptive membrane chromatography for purification of plasmid DNA. J Chromatogr A 989 (2003) 165-73.

[101]A.G. Grunwald, and M.S. Shields, Plasmid purification using membrane-based anion-exchange chromatography. Anal Biochem 296 (2001) 138-41.

[102]D. Kendall, G.J. Lye, and M.S. Levy, Purification of plasmid DNA by an integrated operation comprising tangential flow filtration and nitrocellulose adsorption. Biotechnol Bioeng 79 (2002) 816-22.

[103]D.W. Kahn, M.D. Butler, D.L. Cohen, M. Gordon, J.W. Kahn, and M.E. Winkler, Purification of plasmid DNA by tangential flow filtration. Biotechnol Bioeng 69 (2000) 101-6.

[104]M.D. Butler, D.L. Cohen, D. Kahn, and M.E. Winkler, Patent ID: WO0107599 (2001).

[105]S. Nochumson, Y. Yang, and J.L. Kinsey, Patent ID: WO0194573 (2001).

[106]J.C. Murphy, G.E. Fox, and R.C. Willson, Enhancement of anion-exchange chromatography of DNA using compaction agents. J Chromatogr A 984 (2003) 215-21.

[107]H. GE, PlasmidSelect Xtra for downstream processing of supercoiled plasmid DNA. Application Note 28 (2008) 4094-85.

[108]R. Lemmens, T. Nyhammar, J. Berglof, and J. Stadler, Patent ID: US20056916919. (2005).

[109]P.Å. Albertsson, Partition of cell particles and macromolecules : separation and purification of biomolecules, cell organelles, membranes, and cells in

aqueous polymer two-phase systems and their use in biochemical analysis and biotechnology, Wiley, New York, 1986.

[110]H.-K. Rajni, Aqueous Two-Phase Systems: Methods and Protocols (Methods in Biotechnology), Humana Press, Inc, Totowa, New Jersey, 2000.

[111]K.D. Cole, Purification of plasmid and high molecular mass DNA using PEG-salt two-phase extraction. Biotechniques 11 (1991) 18, 20, 22-4.

[112]R. Ohlsson, C.C. Hentschel, and J.G. Williams, A rapid method for the isolation of circular DNA using an aqueous two-phase partition system. Nucleic Acids Res. 5 (1978) 583-90.

[113]M. Rito-Palomares, and A. Lyddiatt, Impact of cell disruption and polymer recycling upon aqueous two-phase processes for protein recovery. J Chromatogr B Biomed Appl 680 (1996) 81-9.

[114]A. Frerix, M. Muller, M.R. Kula, and J. Hubbuch, Scalable recovery of plasmid DNA based on aqueous two-phase separation. Biotechnol. Appl. Biochem. 42 (2005) 57-66.

[115]B. Zaslavsky, Aqueous Two-Phase Partitioning: Physical Chemistry and Bioanalytical Applications, Marcel Dekker, New York, 1995.

[116]M. Rito-Palomares, Practical application of aqueous two-phase partition to process development for the recovery of biological products. J Chromatogr B Analyt Technol Biomed Life Sci 807 (2004) 3-11.

[117]S. Freitas, S. Canario, J.A. Santos, and D.M. Prazeres, Alternatives for the intermediate recovery of plasmid DNA: performance, economic viability and environmental impact. Biotechnol J 4 (2009) 265-78.

[118]K. Naganagouda, and V.H. Mulimani, Aqueous two-phase extraction (ATPE): an attractive and economically viable technology
for downstream processing of Aspergillus oryzae a-galactosidase. Process Biochem 43 (2008) 1293–1299.

[119]S.K. Sikdar, K.D. Cole, R.M. Stewart, D.C. Szlag, P. Todd, and H. Cabezas, Aqueous two-phase extractions in bioseparations: an assessment. Bio/Technol. 9 (1991) 253-256.

[120]P.A. Rosa, A.M. Azevedo, I.F. Ferreira, J. de Vries, R. Korporaal, H.J. Verhoef, T.J. Visser, and M.R. Aires-Barros, Affinity partitioning of human antibodies in aqueous two-phase systems. J Chromatogr A 1162 (2007) 103-13.

[121]M.E. da Silva, and T.T. Franco, Purification of soybean peroxidase (Glycine max) by metal affinity partitioning in aqueous two-phase systems. J Chromatogr B Biomed Sci Appl 743 (2000) 287-94.

[122]L.J. Karr, J.M. Van Alstine, R.S. Snyder, S.G. Shafer, and J.M. Harris, Cell separation by immunoaffinity partitioning with polyethylene glycol-modified protein A in aqueous polymer two-phase systems. J Chromatogr 442 (1988) 219-27.

[123]L.J. Karr, D.L. Donnelly, A. Kozlowski, and J.M. Harris, Use of poly(ethylene glycol)-modified antibody in cell extraction. Methods Enzymol. 228 (1984) 377-90.

[124]A.F. Buckmanm, M. Morr, and M.R. Kula, Preparation of technical grade polyethylene glycol (Mr 20.000)-N6-(2-aminoethyl)-NADH by a procedure adaptable to large-scale synthesis. Biotechnol. Appl. Biochem. 9 (1987) 258-268.

[125]L.O. Person, and B. Olde, Synthesis of ATP-polyethylene glycol and ATP-dextran and their use in the purification of phosphoglycerate kinase from spinach chloroplasts using affinity partitioning. J. Chromatogr. 457 (1988) 183-193.

[126]G. Birkenmeier, V.P. Shanbhag, and G. Kopperschläger, Comparison of the binding of poly(ethylene glycol)-bound fatty acids to human albumin and alpha-fetoprotein studied by affinity phase partitioning. Biomed Biochim Acta. 45 (1986) 285-9.

[127]P.J. Flory, Principles of Polymer Chemistry, Cornell University Press, Ithaca, NY, 1953.

[128]S.P. Duarte, A.G. Fortes, D.M. Prazeres, and J.C. Marcos, Preparation of plasmid DNA polyplexes from alkaline lysates by a two-step aqueous two-phase extraction process. J Chromatogr A 1164 (2007) 105-12.

[129]B.A. Andrews, A.S. Schmidt, and J.A. Asenjo, Correlation for the partition behavior of proteins in aqueous two-phase systems: effect of surface hydrophobicity and charge. Biotechnol Bioeng 90 (2005) 380-90.

[130]J. Benavides, and M. Rito-Palomares, Practical experiences from the development of aqueous two-phase processes for the recovery of high value biological products. Chemical Technology & Biotechnology 83 (2007) 133-142.

[131]M.e.a. Bensch, High throughput screening techniques in downstream processing: preparation, characterization and optimization of aqueous two-phase systems. Chem. Eng. Sci. 62 (2007) 2011–2021.

[132]S.C. Ribeiro, G.A. Monteiro, J.M. Cabral, and D.M. Prazeres, Isolation of plasmid DNA from cell lysates by aqueous two-phase systems. Biotechnol. Bioeng. 78 (2002) 376-84.

[133]F. Rahimpour, F. Feyzi, S. Maghsoudi, and R. Hatti-Kaul, Purification of plasmid DNA with polymer-salt aqueous two-phase system: optimization using response surface methodology. Biotechnol Bioeng 95 (2006) 627-37.

[134]G.A. Gomes, A.M. Azevedo, M.R. Aires-Barros, and D.M. Prazeres, Purification of plasmid DNA with aqueous two phase systems of PEG 600 and sodium citrate/ammonium sulfate. Sep. Purif. Technol. in press (2008).

[135]C. Kepka, J. Rhodin, R. Lemmens, F. Tjerneld, and P.E. Gustavsson, Extraction of plasmid DNA from Escherichia coli cell lysate in a thermoseparating aqueous two-phase system. J. Chromatogr. A 1024 (2004) 95-104.

[136]A. Frerix, P. Geilenkirchen, M. Muller, M.R. Kula, and J. Hubbuch, Separation of genomic DNA, RNA, and open circular plasmid DNA from supercoiled plasmid DNA by combining denaturation, selective renaturation and aqueous two-phase extraction. Biotechnol Bioeng 96 (2007) 57-66.

[137]C. Kepka, R. Lemmens, J. Vasi, T. Nyhammar, and P.E. Gustavsson, Integrated process for purification of plasmid DNA using aqueous two-phase systems combined with membrane filtration and lid bead chromatography. J Chromatogr A 1057 (2004) 115-24.

[138]C. Kepka, J. Rhodin, and F. Tjerneld, Patent ID: WO04020629 (2000).

[139]A. Frerix, P. Geilenkirchen, M. Muller, M.R. Kula, and J. Hubbuch, Separation of genomic DNA, RNA, and open circular plasmid DNA from supercoiled plasmid DNA by combining denaturation, selective renaturation and aqueous two-phase extraction. Biotechnol. Bioeng. 96 (2007) 57-66.

[140]M. Müller, J. Hubbuch, A. Frerix, and M.R. Kula, Patent ID: US2006286080 (2006).

[141]N.K. Harakas, Protein purification process engineering. Biospecific affinity chromatography. Bioprocess Technol 18 (1994) 259-316.

[142]P.J. Duerksen-Hughes, M.M. Williamson, and K.D. Wilkinson, Affinity chromatography using protein immobilized via arginine residues: purification of ubiquitin carboxyl-terminal hydrolases. Biochemistry 28 (1989) 8530-6.

[143]J. Porath, Immobilized metal ion affinity chromatography. Protein Expr Purif 3 (1992) 263-81.

[144]C.A. Ku, J.D. Henry, Jr., and J.B. Blair, Affinity-Specific protein separations using ligand-coupled particles in aqueous two-phase systems: II. Recovery and purification of pyruvate kinase and alcohol dehydrogenase from Saccharomyces cerevisiae. Biotechnol Bioeng 33 (1989) 1089-97.

[145]U. Sivars, J. Abramson, S. Iwata, and F. Tjerneld, Affinity partitioning of a poly(histidine)-tagged integral membrane protein, cytochrome bo3 ubiquinol

oxidase, in a detergent--polymer aqueous two-phase system containing metal-chelating polymer. J Chromatogr B Biomed Sci Appl 743 (2000) 307-16.

[146]G. Birkenmeier, M.A. Vijayalakshmi, T. Stigbrand, and G. Kopperschlager, Immobilized metal ion affinity partitioning, a method combining metal-protein interaction and partitioning of proteins in aqueous two-phase systems. J Chromatogr 539 (1991) 267-77.

[147]C.R. Lowe, A.R. Lowe, and G. Gupta, New developments in affinity chromatography with potential application in the production of biopharmaceuticals. J Biochem Biophys Methods 49 (2001) 561-74.

[148]A. Klug, Zinc finger peptides for the regulation of gene expression. J. Mol. Biol. 293 (1999) 215-8.

[149]N.P. Pavletich, and C.O. Pabo, Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science 252 (1991) 809-17.

[150]J.H. Laity, B.M. Lee, and P.E. Wright, Zinc finger proteins: new insights into structural and functional diversity. Curr. Opin. Struct. Biol. 11 (2001) 39-46.

[151]J.R. Desjarlais, and J.M. Berg, Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. Proc. Natl. Acad. Sci. USA 90 (1993) 2256-60.

[152]S. Ghose, G.M. Forde, and N.K. Slater, Affinity adsorption of plasmid DNA. Biotechnol Prog 20 (2004) 841-50.

[153]S.B. Stanley, and Crothers, D.M., Lac repressor is a transient gene-activating protein. Cell 51 (1987) 699-707.

[154]M.G. Fried, and Crothers, D.M., Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide-gel electrophoresis. Nucleic Acids Research 9 (1981) 6505-6525.

[155]M. Pfahl, Characteristics of tight-binding repressors of the lac operon. Journal of Molecular Biology 147 (1981) 1-10.

[156]G.M. Forde, S. Ghose, N.K. Slater, A.V. Hine, R.A. Darby, and A.G. Hitchcock, LacO-LacI interaction in affinity adsorption of plasmid DNA. Biotechnol Bioeng 95 (2006) 67-75.

Chapter 2

Selection of Aqueous Two-phase Systems for Affinity Purification of Plasmid DNA

The results presented in this chapter were submitted for publication as follow:

- H.S.C. Barbosa, C. Esteves-Pinto, N.K.H. Slater and J.C. Marcos, Selection of Aqueous Two-phase Systems for Affinity Purification of Plasmid DNA (submitted to Process Biochemistry Journal)

The results described in this chapter were presented in the following conferences:

- H.S.C. Barbosa, N.K.H. Slater and J. Marcos, Selection of Suitable Aqueous Two-Phase System for Plasmid DNA Affinity Purification, National Congress of Biochemistry, Portugal, 2008.
- H.S.C. Barbosa and J.C. Marcos, Plasmid DNA Purification in Poly (ethyleneglycol) / Dextran systems, International Conference on Biopartitioning and Purification, Holland, 2005

2.1 Abstract

Aqueous Two-Phase Systems (ATPSs) used for affinity partitioning of plasmid DNA (pDNA) require that most contaminants accumulate in one phase, together with the pDNA, while the affinity ligand and its complex with pDNA should accumulate in the other phase. In this way the affinity ligand steers the pDNA to a phase with less contaminants, so increasing its purity. In this work the partition of total protein, DNA and RNA from an alkaline bacterial cell lysate in ATPSs composed of poly(ethylene glycol) (PEG) / Dextran (DEX) was examined to identify appropriate systems for pDNA purification. In systems composed by PEG 600 – DEX 40 all major pDNA contaminants accumulated predominantly in the DEX rich phase (protein >80 % and RNA >99 %). These systems are suitable for utilization of DNA affinity ligands based on PEGylated affinity proteins, as well as non-protein ligands that accumulate preferentially in the PEG phase. Total pDNA recoveries in these systems was >80 %, indicating that pDNA precipitation at the interface is low and thus showing them to be appropriate for pDNA affinity purification.

.

Keywords: Aqueous two-phase; Affinity ligands; Poly(ethylene glycol); Dextran; Plasmid purification, Contaminants partitioning

2.2 Introduction

Some of the most promising advances in molecular medicine have been reported in the field of gene therapy and DNA vaccines. Non-viral vectors, such as plasmid DNA gene delivery systems, are been considered for this purpose with great acceptance [1; 2]. However the widespread application of this methodology will require adequate methods to produce and purify plasmids at large scale. Although several chromatographic methods had been proposed for plasmid DNA (pDNA) downstream processing [3; 4], aqueous two-phase systems (ATPSs) represent an attractive alternative.

ATPSs can be scaled-up easily and integrate clarification, concentration and purification in a single operation. Additionally, phase equilibrium is usually achieved in a short time, enabling fast separations that can be performed in continuous mode at low cost [5; 6]. Despite these advantages, ATPSs are not widely adopted in industry, which may be because the prediction of partition behaviour is difficult and suitable systems for a given application are usually found by trial and error. Also, selectivity of ATPSs is usually low and further steps are commonly needed to remove remaining impurities [7; 8; 9].

The use of affinity ligands to enhance the partitioning of a target molecule to a particular phase minimises these two problems by increasing selectivity and makes partition behaviour more predictable [10; 11]. However, such studies have been mainly restricted to the purification of proteins. So far as we know, only our group had reported the utilisation of affinity ligands to enhance plasmid DNA isolation in ATPSs [12; 13].

Efficient utilisation of DNA affinity ligands requires that most contaminants accumulate in one phase, while the affinity ligand and ligand-pDNA complex should accumulate in the opposite phase. In this way the affinity ligand steers the partitioning of pDNA to a contaminant-free phase, so increasing its purity. The selection of appropriate systems to conduct the affinity partition is thus of paramount importance for the success of the process. In our previous studies polymer-polymer systems namely poly(ethylene glycol)(PEG)-dextran were

used. They were preferred to polymer-salt systems since their high salt concentration may impair the affinity binding between the ligand and target pDNA. The composition of each system was selected from preliminary data obtained in our laboratory.

However the further utilization of affinity ligands requires a complete understanding of the partition of the contaminants on the systems and the influence that parameters like polymers molecular weight and concentration might have. Although there is some data about the partition of DNA in literature [5; 14] a systematic study with plasmid DNA from a bacterial lysate in PEG-dextran has not been made yet.

The goal of this work was to study the partitioning of total DNA (DNA_T), total protein ($Protein_T$) and total RNA (RNA_T) from a bacterial cell alkaline lysate in PEG – dextran systems with different polymer molecular weights. Combinations of PEG MW ranging from 600 to 10,000 Da and dextran MW from 40 to 500 kDa were used. For each combination of the phase forming polymers three systems were assessed with different polymer concentrations.

From the results obtained we propose two ATPSs for the efficient utilisation and exploitation of PEGylated and non-PEGylated pDNA affinity ligands in pDNA downstream processing.

2.3 Materials and Methods

2.3.1 Chemicals (Materials)

Poly(ethylene glycol) of molecular weight 600 Da (PEG 600), 1,000 Da (PEG 1000), 3,350 Da (PEG 3350) and 10,000 Da (PEG 10000) was purchased from Sigma (St. Louis, MO, USA). Dextran of molecular weight 40 kDa (DEX 40), 100 kDa (DEX 100) and 500 kDa (DEX 500) was purchased from GE Healthcare (Chalfont St Giles, UK). LB broth media was purchased from Sigma.

2.3.2 Cell Culture

Escherichia coli DH5-α strain, harbouring the pUC19 plasmid was cultivated with solid LB agar media (50 μg/mL Ampicillin) overnight at 37 ºC. One single colony was then used for the inoculation of 10 mL liquid LB media and growth was conducted for 8 h at 37 ºC. 50 μl of this culture was then used to inoculate 200 mL of similar media and growth was conducted overnight at 37 ºC with rotary shaking at 250 rpm.

2.3.3 Preparation of the alkaline lysate

An overnight cell culture was harvested by centrifugation at 6000 *g* for 15 min at 4 ºC and the pellet was resuspended in 8 mL of solution S1 (50 mM glucose, 25 mM Tris-HCl, 10 mM EDTA, pH 8.0). Alkaline lysis was performed by adding 8 mL of solution S2 (200mM NaOH, 1% (w/v) sodium dodecyl sulfate (SDS)) with gentle mixing. The mixture was incubated at room temperature for 10 min. The lysate was then neutralised by adding 8 mL of an ice cold solution S3 (3 M potassium acetate, 11.5% (v/v) glacial acetic acid) and kept on ice for 15 min. The white precipitate containing cell debris, protein and genomic DNA (gDNA) was removed by centrifugation at 15,000 g for 30 min. The supernatant was transferred to a clean tube and a further centrifugation at the same speed for 15 min was used to further clarify the supernatant. The clarified lysate was filtered using a 0.22 μm syringe filter and stored at -20 ºC.

2.3.4 ATPS – Tie Line Determination

Tie-lines were determined by calculating the polymer composition of both the top and bottom phases for each ATPS studied. To each system was added a 20

% (w/w) blank solution of the resulting mixture of solutions S1, S2 and S3, as described above for the bacterial cell lysate preparation. The PEG concentration in both phases was determined by refractometry using a calibration curve constructed with PEG standards of known concentration. The dextran concentration was quantified by polarimetry using calibration curves obtained for dextran solutions of known concentration.

Tie-Line lengths (*TLL*) characterise the compositional differences between the two phases and were calculated using the following formula (1):

$$TLL = \sqrt{(\Delta[PEG])^2 + (\Delta[DEX])^2} \quad (1)$$

where

$$\Delta[PEG] = [PEG]_{top} - [PEG]_{bottom} \;\; ; \;\; \Delta[DEX] = [DEX]_{top} - [DEX]_{bottom} \quad (2)$$

All aqueous two-phase systems compositions are expressed in % (w/w).

2.3.5 ATPS – Preparation

PEG – dextran ATPSs were prepared by weighing appropriate amounts of stock solutions of the phase forming polymers of different molecular weights in 2 mL graduated tubes. A 20 % (w/w) lysate was then added (0.3 g) and the total weight was adjusted to 1.5 g with ultra pure water. The systems were then mixed and centrifuged (2000 g, 2 min) to collect the phases for further analysis. Phase turbidity was avoided by equilibrating the systems at 35 ºC prior to collection of each phase sample. All analyses were conducted at room temperature on the same day that the partitioning systems were prepared. Blank systems were prepared in a similar way but replacing the lysate by 20 % (w/w) of the resulting mixing solution of solutions S1, S2 and S3.

2.3.6 Analytical Methods

2.3.6.1 DNA quantification

Total DNA quantification was performed using a Picogreen assay (Invitrogen, Carlsbad, CA, USA) in a 96 well plate reader. Blanks for each system were prepared and a series of calibration curves using standard concentrations of purified pUC19 plasmid in the blank top or bottom phase of each system was used. To each blank ATPS the lysate was substituted by the mixture of solutions

S1, S2 and S3 with exactly the same final composition as used during the alkaline lysis.

2.3.6.2 RNA quantification

The amount of RNA in each phase was quantified using a nucleic acid-binding reagent, Ribogreen (Invitrogen). Calibration curves were prepared in a similar way to the DNA assay but using pure RNA (from Ribogreen Kit). The contribution of fluorescence arising from DNA binding to Ribogreen was eliminated according to the manufacturer´s manual. Briefly, DNA was digested by incubating the samples for more then 90 min at 37 ºC with RNAse free DNAse (Sigma, 5 units per microgram DNA presumed in the sample) in 0.11 volume of 10X DNAse digestion buffer (200mM Tris-HCl, pH 7.5, containing 100 mM $MgCl_2$ and 20 mM $CaCl_2$). Each sample was diluted 120 times to diminish the effects of salts and polymers present. Fluorescence was measured at 500 nm excitation wavelength and 525 nm emission wavelength in the fluorescence spectrometer.

2.3.6.3 Protein Quantification

Quantitative determination of the protein content in each phase was performed by Bradford assay using BSA as a standard. The interference of PEG and DEX was overcome by diluting the samples ten-fold and making a series of calibration curves using the diluted phase under analysis as solvent for the BSA standards.

2.3.7 Analysis Parameters

The partition coefficient (*Kx*) is the ratio of the concentration in each phase, defined as:

$$Kx = \frac{[x]_{top}}{[x]_{bottom}} \qquad (3)$$

Where $[x]$ is the concentration of the each component x in the respective phase.

Total Recovery (R_T) of DNA is defined as the ratio of DNA (mass) in the top phase plus the bottom phase to the mass of the DNA added, DNA_i:

$$R_T = \frac{[DNA]_{top} \times V_{top} + [DNA]_{bottom} \times V_{bottom}}{DNA_i} \times 100 \quad (4)$$

Where V is the phase volume. Phase recovery (*PR*) of each component x in each phase (i) from the total recovered component was calculated as:

$$PR = \frac{[x]_i \times V_i}{x_{top} + x_{bottom}} \quad (5)$$

2.4. Results and Discussion

2.4.1 Aqueous two-phase systems

The partitioning of total DNA, RNA and protein (DNA_T, RNA_T and $Protein_T$) from an alkaline bacterial cell lysate (20 % w/w) harbouring the pUC19 plasmid was studied in ATPSs composed by different molecular weight combinations of PEG and dextran (Fig. 1 inset). PEG molecular weights ranging from 600 to 10,000 Da and Dextran from 40 to 500 kDa were used. For each combination of the phase forming polymers three systems were assessed with different tie-line lengths (Fig. 1). These systems were named short TLL, medium TLL and long TLL systems according to the relative size of the tie-line. The TLL in any given system is determined by the total polymer concentration in each phase and is a measure of the relative difference between the phases composition (see equation 1).

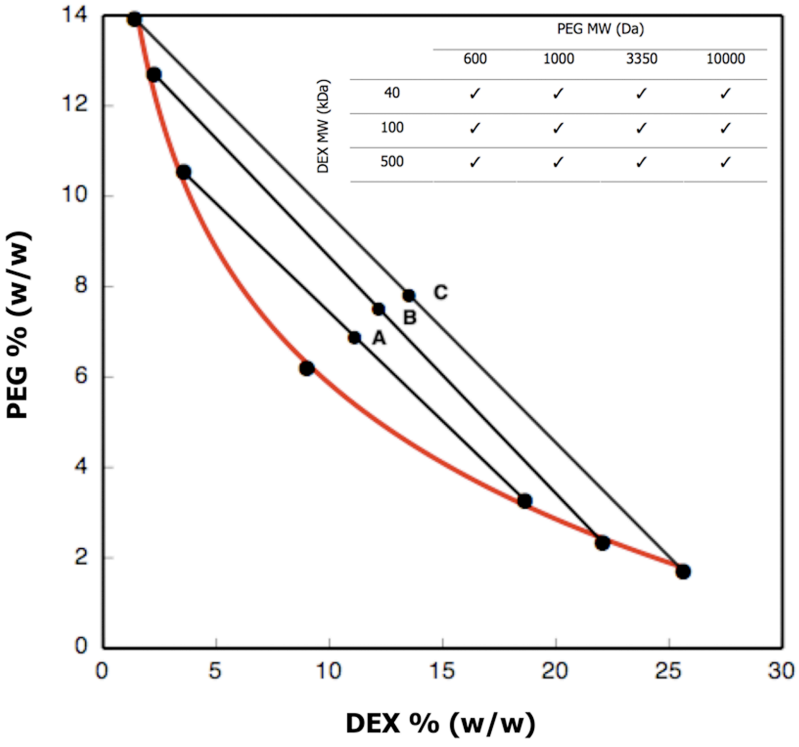


Figure 1 - Example of phase diagram showing the three systems with increased TLL (A- short TLL system, B – medium TLL system and C – long TLL system) selected per each combination of different MW polymers used. The binodal represented is for the system composed by PEG 3350 – DEX 40. Inset table shows the combinations of the molecular weights (MW) selected of the phase forming polymers for each ATPS.

As TLL decreases and approaches zero (at the critical point) the polymer composition in each phase differentiates less from the opposite phase and partition coefficients (K) approach one. Consequently, in this study systems with TLL > 8 % (w/w) (Table 1) were chosen to distinguish from systems where $K \approx 1$. This allows the selection of ATPSs where biomolecules accumulate preferentially in one of the phases.

2.4.2 Partition studies

Fig. 2 shows the variation of the partition coefficient logarithm (log K) of DNA_T, RNA_T and $Protein_T$ as a function of TLL for the different PEG/dextran compositions. In the 36 systems analysed the log K values for total DNA and RNA were always lower then zero independently of the molecular weight of the phase forming polymers and the TLL of the system. This means that both DNA and RNA accumulated in the bottom phase. Similarly the total protein from these bacterial cell lysates accumulated predominantly in the bottom phase of the ATPSs, but its partition coefficient was always higher then the partition coefficient of DNA or RNA in the same system. The logarithm of the total protein partition coefficient varied from -0.52 to 0.33 within all the systems whereas for DNA this parameter ranged from -1.9 and -1.1 and for RNA between total accumulation in the bottom phase and -1.1.

Proteins preferentially accumulated in the bottom phase of most systems with the only exception being for the ones composed of PEG 1000 – DEX 500, where log $K_{Protein}$ was 1.6, 1.8, and 2.2 for short, medium and long TLL length's systems, respectively. Similar systems made up of PEG 1000, but in which DEX 500 was replaced by DEX 40 or by the DEX 100, did not show log $K_{Protein}$ values above 1 and thus the total protein no longer accumulated in the top phase if lower molecular weight dextrans were used. This distinctiveness of DEX 500 systems is usually attributed to the excluded volume effect (as MW of the phase forming polymers increases, less protein accumulates in the same phase due to exclusion) and has been observed for several proteins [15; 16].

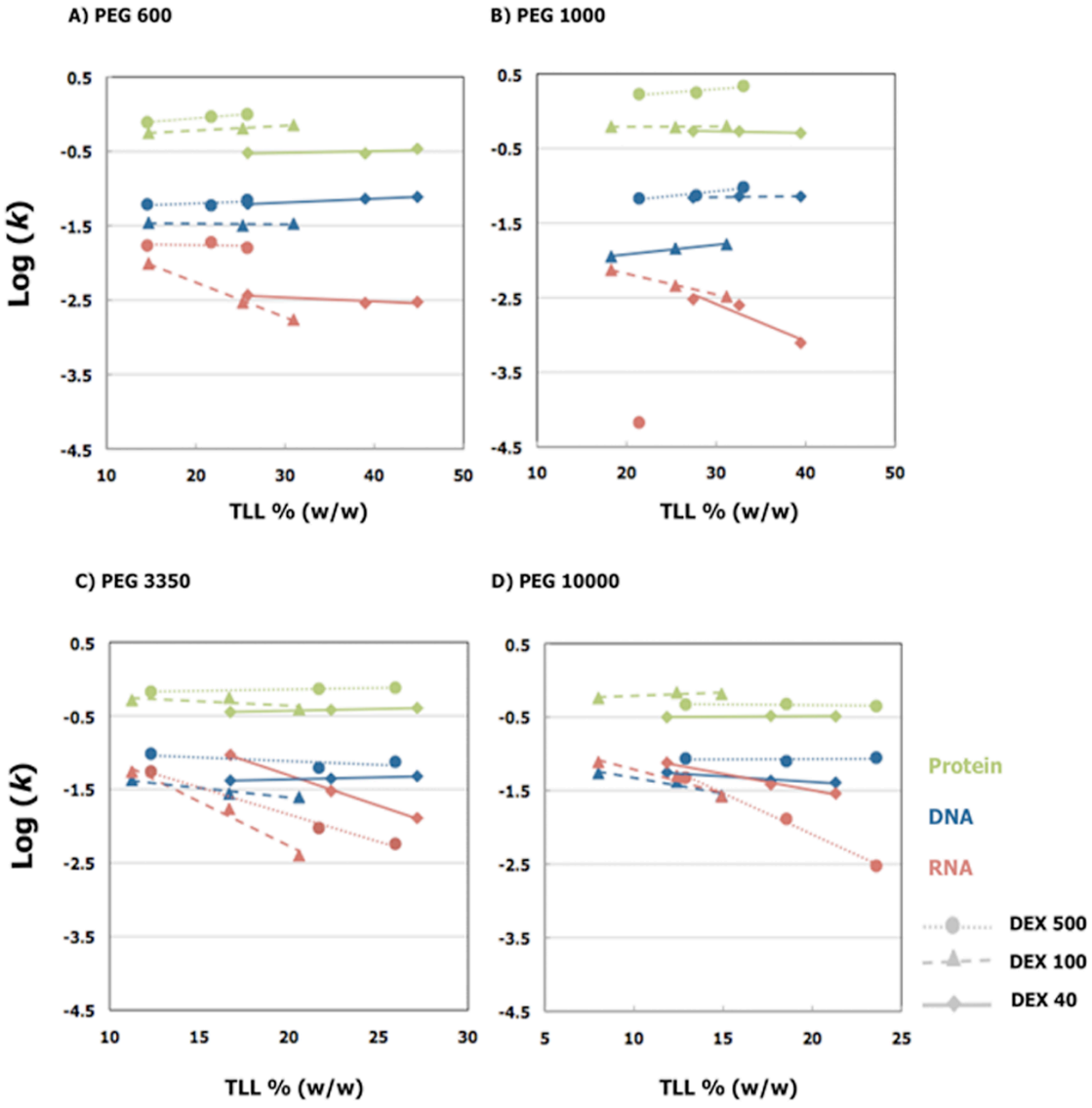


Figure 2 - Partitioning coefficient (K) of total Protein, total DNA, and total RNA in systems composed by poly (ethylene glycol) - dextran with different molecular weights. A) **Systems composed by PEG 600 – DEX (500; 100, 40);** B) **PEG 1000 – DEX (500; 100, 40) systems;** C) **PEG 3350 – DEX (500; 100, 40) systems;** D) **PEG 10000 – DEX (500; 100, 40) systems.**

The same effect was not readily observed for DNA and RNA partitioning and this could be due to their electrostatic charge and size. Although there are a limited number of studies of the partitioning of nucleic acids in polymer/polymer ATPSs, the available data shows that partition behaviour is extremely sensitive to ionic composition. In the initial work by Albertsson [5] it was shown that the partitioning of DNA in PEG 6000 - DEX 500 systems with

10 mM sodium phosphate buffer at pH 6.8 was extremely dependent on the addition of small amounts of salts. For instance the addition of 5 mM NaCl decreased the partition coefficient of DNA from 10 to 2 whilst with 10 mM NaCl the partition coefficient fell to 0.2. Similar effects were observed with other salts, such as RbCl and CsCl. Other have similarly reported that in DEX 500/breox systems, a shift in plasmid DNA partitioning and accumulation from the top to the bottom phase was observed upon increasing the concentration of sodium phosphate from 50 mM to 100 mM [17].

In the present study the precise concentration of salts is unknown as the lysate solution resulted from the mixture of solutions S1, S2 and S3 (see Materials and Methods) and some precipitation occurred. However, given the high concentration of potassium acetate and acetic acid in solution S3 it is expected that the final acetate concentration would have been in excess of 200 mM. Potassium ion concentration would have been of a similar magnitude and this high salt concentration may justify the observed partition behaviour.

Total DNA accumulated preferentially in the bottom phase (> 91.3 %) in all the ATPSs studied, as shown in Table 1**.** No genomic DNA was detected by agarose gel analysis in either the top or bottom phases of these systems (Fig. 3) indicating that gDNA had been removed during the alkaline cell lysis process. The RNA accumulated more unevenly in all systems with the highest partition coefficient of log K_{RNA} = -1.03 obtained for the PEG 3350 – DEX 40 system. This signifies that more than 92 % of the recovered RNA accumulated in the dextran phase.

In systems composed of PEG 1000 – DEX 500 the RNA was found to accumulate exclusively in the bottom phase and thus complete removal of this contaminant in pDNA downstream processing could be achieved in one extraction step.

Table 1 - Systems composition and bottom phase recovery for total Protein, total DNA and total RNA in all ATPS analysed

ATPS (nº)	System Composition				TLL (%, w/w)	Phase ratio (Ø)	Bottom Phase Recovery %		
	PEG MW (Da)	DEX MW (x10^3 Da)	PEG (WT, %)	DEX (WT, %)			Protein	DNA	RNA
1	600		17.5	16.0	25.80	1.07	75.53 ± 3.29	93.80 ± 8.36	99.60 ± 3.09
2	600		18.5	19.0	38.99	0.86	80.78 ± 2.08	94.12 ± 8.78	99.75 ± 1.79
3	600		19.0	21.0	44.82	0.75	79.60 ± 4.57	94.43 ± 3.04	99.77 ± 1.14
4	1000		13.6	14.0	27.43	1.00	64.92 ± 0.77	93.52 ± 8.24	99.69 ± 3.87
5	1000		14.3	16.0	32.58	1.07	64.77 ± 1.10	92.77 ± 6.98	99.73 ± 2.88
6	1000	40	16.0	19.0	39.44	1.16	64.10 ± 0.78	92.25 ± 1.18	99.90 ± 2.81
7	3350		7.0	11.0	16.73	0.87	76.22 ± 4.74	96.48 ± 7.60	92.46 ± 8.25
8	3350		7.5	12.0	22.35	1.08	70.84 ± 9.53	95.42 ± 5.29	97.86 ± 8.42
9	3350		8.0	13.5	27.16	1.00	71.31 ± 1.09	95.44 ± 3.94	98.72 ± 1.42
10	10000		4.0	8.5	11.85	0.78	80.37 ± 3.34	95.84 ± 6.84	94.77 ± 2.40
11	10000		4.6	10.0	17.67	0.75	79.74 ± 3.84	96.77 ± 2.94	97.11 ± 2.32
12	10000		5.5	11.0	21.31	0.75	79.97 ± 2.28	96.95 ± 3.01	97.81 ± 0.86
13	600		17.6	13.7	14.69	0.81	68.94 ± 4.92	97.30 ± 2.72	99.21 ± 3.82
14	600		15.9	16.2	25.26	0.85	65.18 ± 2.31	97.43 ± 2.41	99.84 ± 1.82
15	600		16.2	17.4	30.94	0.85	62.81 ± 1.79	97.30 ± 3.98	99.86 ± 1.13
16	1000		12.2	14.4	18.27	1.10	58.44 ± 6.98	98.72 ± 0.87	99.15 ± 1.05
17	1000		12.5	16.3	25.45	0.86	64.17 ± 4.54	98.69 ± 3.07	99.58 ± 2.02
18	1000	100	13.0	17.7	31.17	0.87	66.02 ± 4.02	98.68 ± 5.39	99.74 ± 4.06
19	3350		6.1	10.6	11.24	0.66	76.05 ± 4.53	97.46 ± 0.98	96.74 ± 1.70
20	3350		6.4	11.7	16.66	0.75	70.48 ± 4.67	97.95 ± 5.55	98.72 ± 0.70
21	3350		6.6	12.7	20.57	0.86	74.19 ± 8.34	97.83 ± 6.97	99.64 ± 0.78
22	10000		3.9	8.07	8.21	0.73	69.75 ± 4.25	96.03 ± 2.14	94.48 ± 0.95
23	10000		3.1	8.82	12.39	0.75	65.93 ± 4.90	96.96 ± 5.54	96.33 ± 1.02
24	10000		3.6	9.61	14.91	0.93	62.10 ± 1.62	97.54 ± 1.65	97.59 ± 2.01
25	600		16.0	9.0	14.54	1.14	57.92 ± 3.71	94.85 ± 7.57	98.41 ± 1.77
26	600		16.5	11.0	21.70	0.75	51.95 ± 4.73	94.38 ± 5.45	98.13 ± 0.40
27	600		16.8	12.5	25.75	1.00	50.19 ± 6.78	93.42 ± 5.18	98.42 ± 3.43
28	1000		12.5	11.0	21.38	0.86	40.51 ± 7.54	94.34 ± 5.14	99.99 ± 0.41
29	1000		13.0	13.0	27.78	1.00	36.07 ± 0.98	93.27 ± 2.14	100.18 ± 1.99
30	1000	500	13.5	15.5	33.04	1.00	31.45 ± 3.39	91.28 ± 6.37	100.17 ± 1.24
31	3350		6.0	7.0	12.30	0.55	72.78 ± 5.08	94.90 ± 1.10	97.00 ± 0.16
32	3350		6.5	10.5	21.67	0.86	61.01 ± 1.46	94.87 ± 1.98	99.18 ± 1.25
33	3350		7.0	12.0	25.93	0.75	63.62 ± 8.35	94.66 ± 3.77	99.56 ± 1.77
34	10000		3.6	6.0	12.89	0.93	69.59 ± 1.20	92.37 ± 2.39	95.74 ± 0.87
35	10000		4.2	9.0	18.55	0.75	73.74 ± 3.40	94.52 ± 4.89	99.03 ± 1.92
36	10000		5.0	11.0	23.59	0.75	74.99 ± 3.60	93.45 ± 6.50	99.77 ± 4.55

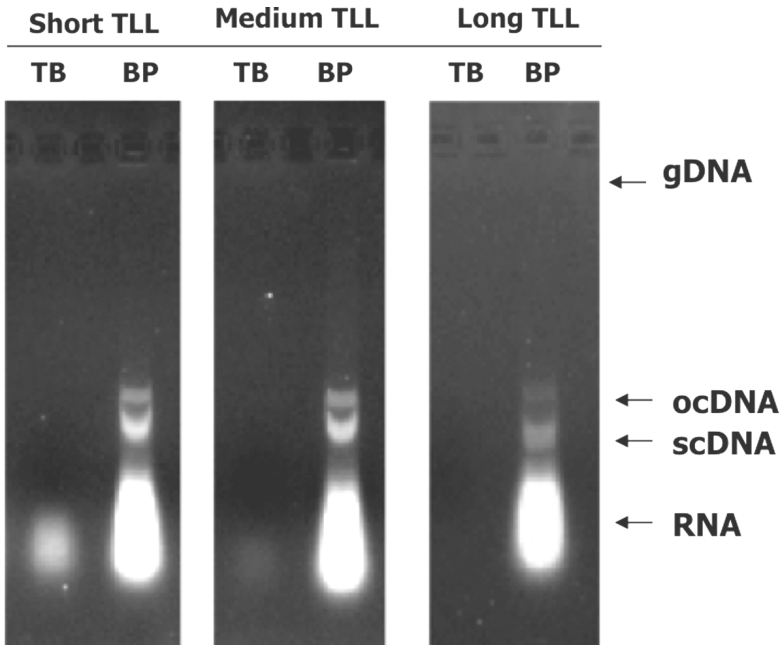


Figure 3 - Agarose gel analysis of the top phase (TP) and Bottom phase (BP) samples of the three systems studied (Short TLL, Medium TLL and Long TLL) for the PEG 10000 – DEX 40 systems. Open Circular plasmid DNA (ocDNA); Supercoiled plasmid DNA (scDNA) and genomic DNA (gDNA).

The increase of TLL in each system had little effect on the protein and pDNA partition coefficients in all of the systems studied (Fig. 2). In contrast, the RNA partition coefficient decreased with increasing TLL (see also agarose gel, Fig. 3). This effect is more evident when high MW PEGs were used. Nevertheless, the percentage of RNA remaining in the top phase was always below 8 % of the total RNA recovered.

Total pDNA recoveries in these polymer – polymer ATPSs were usually above 80 % (Fig. 4), being greater for short TLL systems. Nearly 100 % recovery was achieved using the PEG 3350 – DEX 100 system. Systems composed of low MW PEG and DEX 40 displayed low pDNA recoveries (some systems below 80 %) and the same trend was seen in long TLL systems composed of DEX 500. It is likely that this effect is due to increased interfacial tension between the two phases as the TLL increased, promoting the precipitation of pDNA at the interface. A similar trend has been observed for systems composed by PEG/ammonium sulphate but to a greater extent than observed here, reflecting the higher interfacial tension in polymer/salt systems than in polymer/polymer systems [8].

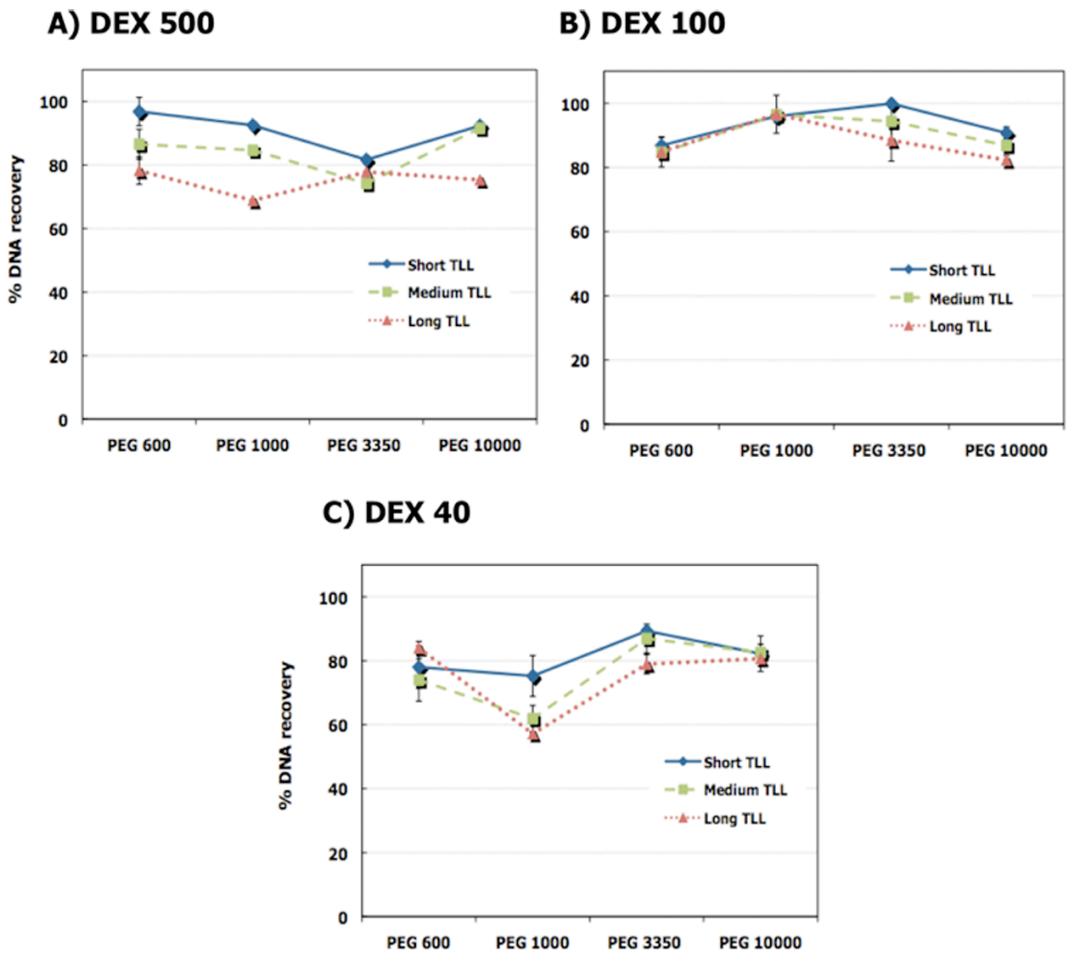


Figure 4 - Total DNA recovery for all the ATPSs systems screened. A) **Systems composed with the DEX 500 – PEG (600; 1000; 3350; 10000);** B) **Systems composed with the DEX 100 – PEG (600; 1000; 3350; 10000);** C) **Systems composed with the DEX 400 – PEG (600; 1000; 3350; 10000).**

2.4.3 ATPS selection for pDNA affinity isolation

From the results of this study we divided the ATPSs into two categories (A and B) that differ mainly in the partitioning of total protein from the bacterial cell lysate. In Category A systems, protein accumulated mainly in the bottom dextran phase together with the pDNA and RNA. By contrast, in category B systems protein accumulated in the PEG rich phase, whilst pDNA and RNA accumulated in the bottom phase.

The systems composed by PEG 1000 – DEX 500 were the only ones to fit into category B and all the other ATPSs could be included in category A. Nevertheless systems composed of DEX 500 showed a high accumulation of proteins in the top phase due to the exclusion effect, as discussed above. For

instance in the systems composed of PEG 600 – DEX 500, only 50.2 % of total protein accumulated in the bottom phase so the protein was almost equally distributed between the two phases.

The system composed by 18.5 % (w/w) PEG 600 – 19.0 % (w/w) DEX 40 was the one to best fit into category A since it offered the highest protein accumulation in the bottom phase (> 80 %) and it also had 94 % of the DNA and 99 % of RNA in the bottom phase, as shown in Table 1.

Both categories A and B could be used in affinity purification processes, although employing different strategies. In category A systems a one-step elimination of most pDNA contaminants is achieved. Nevertheless, chemical modification of the DNA affinity ligand (e.g. PEGylation) might prove to be necessary since is likely that non-modified protein based ligands will accumulate in the bottom phase of these systems as do lysate proteins. Systems within category B may allow the use of protein-based ligands without previous chemical modification, since they might accumulate in the PEG phase, thus reducing operating costs and time compared to systems where the PEGylation is necessary. However, a further extraction may have to be done in order to remove any lysate proteins that are accumulated into the same phase where the protein ligand accumulates.

Although prediction of the partitioning of a particular ligand is difficult, and has to be confirmed experimentally, we have shown recently that the DNA affinity binding protein, zinc finger–GST (Glutathione-S-Transferase) partitions as envisaged here for systems A and B [13]. The non-PEGylated protein was observed to accumulate in the top phase of systems composed by PEG 1000 – DEX 500 in contrast with systems composed by PEG 600 – DEX 40, where the protein accumulated mainly in the bottom phase. When using the PEG 600 – DEX 40 systems we also showed that the zinc finger–GST partitioning could be greatly enhanced by PEGylation. In other preliminary studies with another DNA affinity binding protein (LacI) we have observed the same partitioning trend in each category (data not published).

Table 2 summarises the advantages and drawbacks of using either category A or B systems for the utilization of pDNA affinity ligands in ATPSs. Both category A and B systems achieved good removal of RNA (almost 100 %) in the first extraction. The second extraction used for the removal of remaining impurities could be also used for pDNA elution from the ligand and for the recovery of the affinity ligand. For this step a PEG-salt system is preferred, as a high salt concentration would aid the separation of pDNA from the protein ligand since the interactions between the two are usually electrostatic.

Table 2 - Advantages and drawbacks of using systems A or B for the DNA affinity purification in ATPS

System Category	Protein Ligands		Non-Protein Ligand	Advantages	Drawbacks
	Native protein	PEGylated Protein			
(A) PEG 600 – DEX 40 [a)]	Not Suitable	Good	Good	Elimination of all major contaminants in the first extraction	Protein PEGylation needed Addition partition studies required [b)]
(B) PEG 1000 – DEX 500	Good	Not Suitable	Not Suitable	No need for PEGylated ligand [c)] Total removal of the RNA	High Protein impurities in the first extraction

a) Alternatively, system PEG 3350 – DEX 100 could also be used with similar results.
b) If non-PEGylated protein is used additional ligand partitioning studies could be necessary to find appropriate conditions.
c) If protein-based ligands are used.

A schematic representation of the suggested process is presented in Fig. 5. However, as shown in this study, polymer – polymer ATPSs are preferable to polymer – salt systems for the first extraction system. In polymer-salt systems the partitioning of plasmid DNA and its contaminants could be very dependent on the PEG MW, TLL or salts used [8; 18; 19]. A system where all contaminants accumulated in the bottom phase would be much more difficult to predict as compared to PEG/dextran systems.

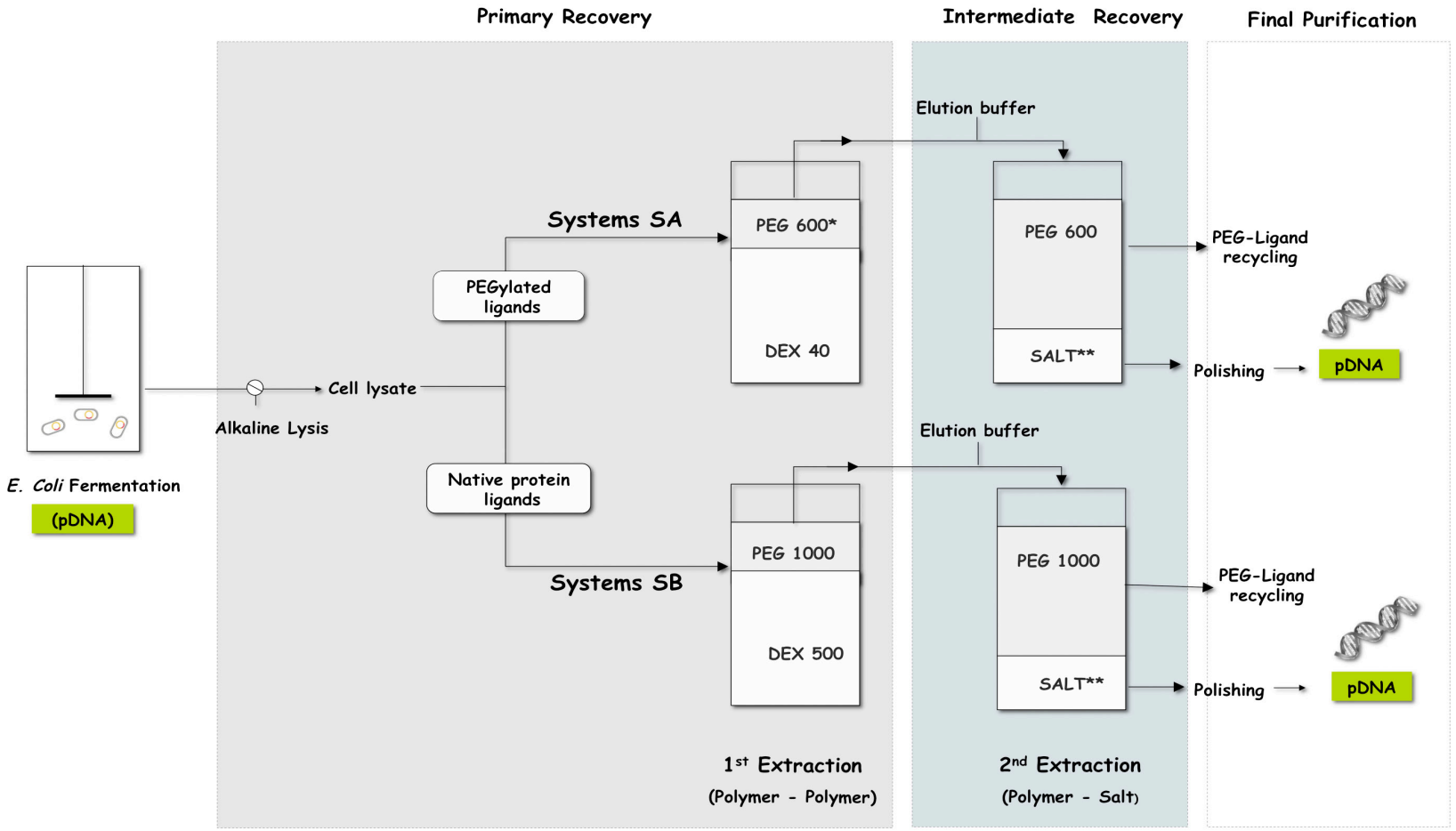


Figure 5 - Schematic representation of the two suitable paths for utilisation of different DNA ligands.
* Alternatively PEG 3350 – DEX 100 could be used;
** Polymer as second phase may be necessary if the salt precipitates the ligands or pDNA.

2.5 Conclusions

We selected two suitable ATPSs for the affinity purification of plasmid DNA in downstream processing applications. Protein or non-protein based DNA affinity ligands might be easily used in the selected systems without the tedious process of finding an optimal system for the efficient removal of pDNA contaminants (gDNA, RNA and Proteins). Systems were selected where all the major pDNA contaminants accumulated in the bottom phase. A removal of more than 80 % of protein, 99 % of the RNA and ~100% genomic DNA can be achieved with only minimal loss of pDNA at the phase interface.
We propose that the utilisation of affinity ligands that naturally accumulate in the top phase, or that can be modified with PEG molecules (PEGylation) to steer their partitioning to this phase, could be best used in category A systems. The category B system may be suitable for the utilisation of non-modified protein affinity ligands.

We have also shown that polymer-polymer ATPSs confer a better choice for the first extraction systems when using pDNA affinity ligands compared to the PEG-salt systems described in the literature since the partitioning of bacterial cell components varies less with different polymer MWs, or TLL of each system, despite the fact that salt might also impair the binding of the ligand to the pDNA. The utilisation of pDNA ligands in ATPSs might therefore improve upon the low selectivity usually observed in these systems compared to the chromatographic methods. Using these system compositions for pDNA purification may enable highly selective isolation of pDNA.

2.6 References

[1]A. Mountain, Gene therapy: the first decade. Trends Biotechnol 18 (2000) 119-28.

[2]G.M. Forde, Rapid-response vaccines--does DNA offer a solution? Nat Biotechnol 23 (2005) 1059-62.

[3]P.A. Shamlou, Scaleable processes for the manufacture of therapeutic quantities of plasmid DNA. Biotechnol Appl Biochem 37 (2003) 207-18.

[4]D.M. Prazeres, G.N. Ferreira, G.A. Monteiro, C.L. Cooney, and J.M. Cabral, Large-scale production of pharmaceutical-grade plasmid DNA for gene therapy: problems and bottlenecks. Trends Biotechnol 17 (1999) 169-74.

[5]P.Å. Albertsson, Partition of cell particles and macromolecules : separation and purification of biomolecules, cell organelles, membranes, and cells in aqueous polymer two-phase systems and their use in biochemical analysis and biotechnology, Wiley, New York, 1986.

[6]R. Hatti-Kaul, Aqueous two-phase systems: A general overview. Appl Biochem Biotech 19 (2001) 269-277.

[7]F. Rahimpour, F. Feyzi, S. Maghsoudi, and R. Hatti-Kaul, Purification of plasmid DNA with polymer-salt aqueous two-phase system: Optimization using response surface methodology. Biotechnol Bioeng 95 (2006) 627-637.

[8]I.P. Trindade, M.M. Diogo, D.M.F. Prazeres, and J.C. Marcos, Purification of plasmid DNA vectors by aqueous two-phase extraction and hydrophobic interaction chromatography. J Chromatogr A 1082 (2005) 176-184.

[9]G.A. Gomes, A.M. Azevedo, M.R. Aires-Barros, and D.M.F. Prazeres, Purification of plasmid DNA with aqueous two phase systems of PEG 600 and sodium citrate/ammonium sulfate. Sep Purif Technol (2008).

[10]G. Johansson, Affinity partitioning of proteins. Methods Mol Biol 147 (2000) 105-17.

[11]B. Zaslavsky, Aqueous Two-Phase Partitioning: Physical Chemistry and Bioanalytical Applications, Marcel Dekker, New York, 1995.

[12]S.P. Duarte, A.G. Fortes, D.M. Prazeres, and J.C. Marcos, Preparation of plasmid DNA polyplexes from alkaline lysates by a two-step aqueous two-phase extraction process. J Chromatogr A 1164 (2007) 105-12.

[13]H. Barbosa, A.V. Hine, S. Brocchini, N.K. Slater, and J.C. Marcos, Affinity partitioning of plasmid DNA with a zinc finger protein. J Chromatogr A O 1206 (2008) 105-12.

[14]R. Ohlsson, C.C. Hentschel, and J.G. Williams, A rapid method for the isolation of circular DNA using an aqueous two-phase partition system. Nucleic Acids Res 5 (1978) 583-90.

[15]N.L.P. Dallora, J.G.D. Klemz, and P.d.A. Pessoa Filho, Partitioning of model proteins in aqueous two-phase systems containing polyethylene glycol and ammonium carbamate. Biochem Eng J 34 (2007) 92-97.

[16]S. Sasakawa, and H. Walter, Partition behavior of native proteins in aqueous dextran-poly(ethylene glycol)-phase systems. Biochemistry 11 (1972) 2760-5.

[17]C. Kepka, J. Rhodin, R. Lemmens, F. Tjerneld, and P.E. Gustavsson, Extraction of plasmid DNA from Escherichia coli cell lysate in a thermoseparating aqueous two-phase system. J Chromatogr A 1024 (2004) 95-104.

[18]S.C. Ribeiro, G.A. Monteiro, J.M.S. Cabral, and D.M.F. Prazeres, Isolation of plasmid DNA from cell lysates by aqueous two-phase systems. Biotechnol Bioeng 78 (2002) 376-384.

[19]A. Frerix, M. Muller, M.R. Kula, and J. Hubbuch, Scalable recovery of plasmid DNA based on aqueous two-phase separation. Biotechnol Appl Biochem 42 (2005) 57-66.

Chapter 3

Modified Fusion Proteins for Affinity Aqueous Extractions:
PEGylation of Glutathione-S-Transferase (GST)

Subchapter 3.1

Affinity Partitioning of Plasmid DNA with a Zinc Finger Protein

Subchapter 3.2

The results presented in this chapter were published or submitted for publication as follow:

- H.S.C. Barbosa, A. V. Hine, S. Brocchini, N.K.S. Slater and J.C. Marcos (2008), Affinity Partitioning of Plasmid DNA with a Zinc Finger Protein. J Chromatogr A, 1206(2): 105-12.

- H.S.C. Barbosa, S. Balan, N.K.H. Slater, S. Brocchini and J.C. Marcos, Modified Fusion Proteins for Affinity Aqueous Extractions: PEGylation of Glutathione-S-Transferase (GST) (submitted to Biotechnlogy Progress Journal)

The results described in this chapter were presented in the following conferences:

- H.S.C. Barbosa *et al.*, PEGylation of GST-ZF protein: a DNA affinity ligand for plasmid DNA purification in Aqueous Two-Phase Systems, Fourth UK Biochemical Engineering Research Showcase, Birmingham, UK (oral presentation)

- H.S.C. Barbosa, *et al.*, Affinity purification studies of pDNA in Aqueous Two-Phase Systems using a PEGylated GST-ZnF protein, International Conference on Biopartitioning and Purification, Lisbon, Portugal (oral presentation)

- H.S.C. Barbosa, *et al.*, PEGylation of GST-ZF protein: a DNA affinity ligand for plasmid DNA purification in aqueous two-phase systems, 6th European Symposium in Biochemical Engineering Science, Austria (oral presentation)

Subchapter 3.1

Modified Fusion Proteins for Affinity Aqueous Extractions:
PEGylation of Glutathione-S-Transferase (GST)

3.1.1 Abstract

Protein-based ligands are commonly used for affinity extractions of biomolecules in aqueous two-phase systems (ATPSs). To increase the effectiveness of the purification process, the affinity ligand can be covalently conjugated to one of the phase forming polymers, usually poly (ethylene glycol) (PEG). This can assist the partitioning of the target biomolecule to a specific phase, which is usually the PEG phase. Here, a comparative PEG conjugation study of Glutathione-*S*-Transferase (GST) and a GST-zinc-finger fusion protein (GST-ZnF) is reported. GST is a common protein-tag used in recombinant applications. The GST-zinc finger fusion protein is designed to bind DNA, hence the purpose of its PEGylation is to increase the selectivity of DNA purification by ATPS.

Four different functionalised PEGylation reagents were examined for GST modification. Two amine selective reagents, 20 kDa mPEG-butyraldehyde (mPEG-ALD) and 20 kDa mPEG-succinimidyl propionate (mPEG-SPA) were evaluated. Two thiol selective reagents, 5 kDa mPEG-maleimide (mPEG-MAL) and the bis-alkylating 10 kDa mPEG-monosulfone (mPEG-MS).

Thiol specific GST PEGylation was achieved in high conversion (approximately quantitative) with both of the thiol selective reagents in about 1h reaction time (20 PEG equivalents (equiv)). The amine specific PEGylation methods were less efficient. PEGylation reactions were then conducted with GST-ZnF and similar efficiencies were observed. All of the PEGylation products displayed increased protein partition coefficients to the PEG rich phase in ATPS comprised of PEG and dextran. The affinity partitioning of a DNA oligonucleotide containing the ZnF recognition site was evaluated for the differently PEGylated GST-ZnF protein and the results presented.

Keywords: Glutathione-*S*-Transferase; zinc-finger protein; PEGylation; plasmid purification; aqueous two-phase systems.

3.1.2 Introduction

The covalent conjugation of poly (ethylene glycol) (PEG) to proteins (PEGylation), first introduced in the 1970's, has been shown to be applicable for a wide range of protein therapeutics. PEGylation increases the pharmacokinetic profile of protein therapeutics by improving their circulation half-life and efficacy [1; 2; 3; 4].

Applications of PEGylation in biopharmaceutical purification are also of interest. The purification of biomolecules using aqueous two-phase systems (ATPS) can be conducted using specific ligands for the target product, which can be covalently bound to one of the phase forming polymers, usually poly (ethylene glycol) (PEG). The utilization of a PEGylated protein ligand can enhance the partitioning of the target product into the PEG-rich phase. If ATPSs are selected in which major impurities accumulate in the opposite phase then this approach can potentially allow a significant improvement in the selectivity and effectiveness of the purification process, resulting in the higher yield and purity of the final product [5]. Several protein-based affinity ligands (Protein A [6], antibodies [7]), or small molecule-ligands (NADH [8], ATP [9], fatty acids [10] and reactive dyes [11]) have been grafted to PEG and used in this way.

A common chromatographic strategy for protein purification is to fuse it to Glutathione - S - Transferase (GST) (EC 2.5.1.18) as an affinity tag. The fusion protein can be expressed in a suitable host strain and recovered by affinity chromatography using immobilized glutathione as ligand that binds specifically to the GST moiety of the fusion protein. Following elution, proteolytic cleavage of GST yields the target protein [12; 13].

Provided the functionality of the GST fusion protein is maintained [14], GST fused proteins ligands can be used for bioaffinity extractions by ATPS. This approach to has been used in this laboratory to isolate plasmid DNA (pDNA) by ATPS [15]. A GST - zinc-finger fusion protein (GST-ZnF) was PEGylated and used as an affinity ligand for the isolation of pDNA containing a zinc-finger recognition site. During ATPS isolation pDNA localised in a dextran (DEX) rich

phase, and its extraction to a PEG rich phase by ATPS was achieved using the PEGylated GST-ZnF.

Here, four PEGylation reagents are considered. Two of these reagents are selective for the PEGylation of free amine groups on proteins and the other two are specific for protein thiols. The aim of this comparative study was to optimize these four distinct reactions for PEG conjugation to GST and to evaluate whether the best reaction conditions could be translated to GST-ZnF.

Since the ligand fused to the GST can also be PEGylated, it is important that the PEG conjugation sites are distinct from the ligand binding site. Furthermore, since most binding sites do not have free thiols it was hypothesised that thiol reactive PEG reagents would be least likely to interfere with. Four different PEGylation reagents were thus evaluated to determine their reaction yield, product heterogeneity and reaction time. The reagents used were: **1** 20kDa mPEG-butyraldehyde (mPEG-ALD); **2** 20kDa mPEG-succinimidyl propionate (mPEG-SPA); **3** 5kDa mPEG-maleimide (mPEG-MAL); **4** 10kDa mPEG-monosulfone (mPEG-MS).

There are 21 potential sites (ε-amine groups of lysines (Lys)) for amine PEGylation in the recombinant GST protein plus the N-terminal amino acid residue (α- amino group of the methionine).

The aldehyde reagent **1** achieves PEG conjugation by reductive amination. This two-step reaction proceeds by the formation of an imine intermediate, which is then reduced with sodium cyanoborohydride ($NaCNBH_3$) at acidic pH (~5) [3; 16; 17]. The terminal amino group often has a slightly lower pKa compared to the ε-amine groups of Lys and so can undergo the imine reaction more favourably. However in practice since the imine reaction is not thermodynamically driven, an excess of PEG aldehyde reagent is often required and this can result in non-specific protein PEGylation.

The active ester reagent **2** with the N-hydroxysuccinimide (NHS) leaving group is generally considered to be the most non-specific of the four reagents. While it

can undergo reaction with the N-terminal residue, it can also undergo acylation conjugation reactions with the ε-amine groups of Lys of the GST. The active ester reagent **2** also undergoes competitive hydrolysis during conjugation (hydrolysis half-life – 16.5 min, pH 8 [17]), so excess is usually required in order to achieve a useful degree of PEGylation.

Reagent **3** and **4** were used for site-specific PEGylation of the four cysteines (Cys) thiols within GST. Analysis of the three-dimensional structure of the GST protein (1GTA, PDB Protein Data Bank) reveals that the four Cys residues may not be cross-linked and thus available for PEGylation. However, the experimental data presented in this study shows that only 2 Cys on GST are easily accessible for PEGylation using non-reducing conditions.

Reagent **3** undergoes alkylation reactions with the sulfhydryl groups to form one stable thioether bond to one sulfhydryl [17], whereas **4** is capable of bis-alkylation by sequential addition-elimination reactions forming the 3-carbon bridge between two sulfhydryls with the PEG molecule attached [18; 19]. The bis-alkylation reagent **4** covalently binds to two thiols, hence it was used the 10 kDa molecular weight instead of a 5 kDa reagent (as reagent **3**) so that the relative mass of the PEG for the two thiols could potentially be compared.

It is shown that optimized PEGylation conditions found for free GST could be applied to GST fused to the ZnF transcript factor protein (GST-ZnF). The zinc finger moiety contributes to more 6 Cys and 13 Lys for potential PEG-conjugation. However, the Cys amino acids on the ZnF moiety are believed to be inaccessible for PEGylation in the non-reducing conditions, since together with histidine amino acids they form coordination site for binding a Zn^{2+} ion (the DNA binding site) [20]. The effect of each PEGylation chemistry on the ZnF DNA binding activity and its partitioning in PEG/dextran ATPS is considered.

3.1.3. Methods and Materials

3.1.3.1 Reagents

PEGylation reagents mPEG-butyraldehyde (mPEG-ALD, molecular weight (MW) 20 kDa) **1**, mPEG-succinimidyl propionate (mPEG-SPA, MW 20 kDa) **2,** mPEG-maleimide (mPEG-MAL, MW 5 kDa) **3** were purchased from Nektar (San Carlos, CA, USA) and mPEG-monosulfone reagent (mPEG-MS, MW 10 kDa) was obtain by synthesis [21]. Complementary synthetic oligonucleotides carrying the zinc finger recognition site 5′ - TTT-TTT-TTT-TGG-GGC-GGC-TTT-TTT-TTT-TT - 3′ and 5′ - AAA-AAA-AAA-AAG-CCG-CCC-CAA-AAA-AAA-AA - 3′ were purchased from Biomers (Ulm, Germany). They were annealed as described by [15]. PEG polymer, MW 10,000 Da was obtained from Sigma-Aldrich (St. Louis, MO, USA) and Dextran MW 500 kDa was purchase from GE Healthcare (Chalfont St Giles, UK).

3.1.3.2 Expression and purification of GST protein

GST protein was expressed from the pGEX-2TK vector (GE Healthcare) in *E. coli* BL21 cells and purified as described previously [22]. The resulting eluate containing GST protein was dialysed at 7 ºC against 50 mM Tris-HCl buffer, 130 mM NaCl buffer pH 7.4 to remove reduced glutathione. The purified protein was stored at -20 ºC with 30 % glycerol. Protein concentration in solution was determined with the Bradford method (Pierce, Rockford, IL, USA) using BSA as standard. GST-ZnF fusion protein was expressed, purified and stored in the same conditions described above for GST protein.

3.1.3.3 GST PEGylation

3.1.3.3.1 Protein amine specific PEGylation of GST: reagent 1 mPEG-ALD

Reductive alkylation was performed by mixing purified GST protein (typically 0.2-0.4 mg/mL) with the PEG aldehyde reagent **1** at varied GST molar equivalents (equiv) of the mPEG (ranging 1 to 50 equiv) in 50 mM sodium acetate buffer (1mL), pH 5.0 (unless otherwise stated). After adding the PEG

solution, sodium cyanoborohydride (50 µL of 60 mg/mL solution) was added to achieve the final concentration of 3 mg /mL of the reducing agent per sample (unless otherwise stated). The mixture was thoroughly mixed and was incubated at 7 ºC in 1.5 mL tubes for different reaction times (see figures legends for details). Reaction times were varied to optimise the extent of PEGylation. The reactions were quenched by adding 0.2 mL of 2.5M Tris-HCl buffer, pH 7.4 to the final concentration of 0.41M.

A PD-10 desalting column (GE Healthcare) was used to remove low MW reaction end-products. The column was first pre-equilibrated with 50 mM Tris-HCl buffer pH 7.4 and then the reaction mixture (1.0 mL) was loaded onto the PD-10 column. The third and fourth fractions (1.0 mL each) were collected. Typically the first two fractions did not contain any PEG modified protein and were discarded. The PD-10 eluate was then loaded into a cation exchange column (Hitrap SP FF 1 mL, GE Healthcare) previously equilibrated with 50 mM sodium acetate buffer pH 5.0. Ion exchange chromatography was used to remove unreacted PEG compounds. PEG-protein conjugate and unreacted protein were eluted with 50 mM sodium acetate buffer containing 1 M NaCl , pH 5.0. The first fraction (1.0 mL) was discarded and the second and third fractions were collected and combined for further analyses. It is anticipated that this step could be avoided since the excess of PEG would not be a major problem in the utilization of PEGylated ligands in ATPS applications due to the fact that systems are composed with PEG as phase-forming polymer. The separation of the unreacted GST protein from the PEGyated GST was performed by size exclusion chromatography (SEC) as described in section 2.4.2. GST-ZnF fusion protein was PEGylated using the reagent **1** in a similar way as used for GST using the optimal conjugation conditions found with GST, i.e. 25 mPEG-ALD equiv was used in 48h reactions at pH 5 (unless otherwise stated).

3.1.3.3.2 Protein amine specific PEGylation of GST- reagent 2 mPEG-SPA

Typically 0.2 mg/mL GST protein in phosphate buffer saline (50 mM PBS, pH 7.8) in a 1 mL reaction volume was used. The reagent **2** mPEG-SPA was

previously dissolved in dimethylsulfoxide (DMSO) and added to the protein solution. The stoichiometry was varied with increased molar equiv (from 10 to 100) to the GST solution (at room temperature) in an effort to optimise PEGylation. The reaction was then allowed to proceed at 7ºC by slow mixing in end-to-end rotor over 24 h periods. The removal of low molecular weight reaction end-products was performed as described for the reagent **1** reactions, i.e. PD-10 column. Cation exchange and size exclusion chromatography procedures were not employed because this PEGylation was not efficient enough to proceed with further studies.

3.1.3.3.3 Thiol Specific Alkylation of GST – reagent 3 mPEG-MAL

GST protein (typically 0.5 mg/mL, 1 mL) was reduced prior to PEGylation with 100 mM dithiothreitol (DTT) (15.4 mg) at room temperature for 30 min. DTT was removed using a PD-10 desalting column that had been pre-equilibrated with 50 mM sodium phosphate buffer, pH 7.8. The total protein in the PD-10 eluate was quantified (by $UV_{280\ nm}$) and 20 PEG equiv (5 molar excess per Cys) of the activated mPEG- MAL reagent (freshly prepared in the same buffer) was added to the 1 mL reaction volume. The reaction was performed for 1 h at 7ºC. The removal of unreacted PEG was not performed by cation exchange column since the mPEG-MAL has low MW (5 KDa) and can be easily separated as an distinct peak in SEC (see Figure 4B). PEGylated protein products were isolated from the reaction mixture using SEC as described below in section 2.4.2. The sample peaks corresponding to the PEGylated protein were collected and protein was concentrated using a Vivaspin 6 concentrator column. The same reactions were performed in non-reducing conditions by removing the first reducing step with DTT and proceeding in a similar way. GST-ZnF was used as reduced and non-reduced form for PEGylated using this reagent in the same stoichiometry (20 PEG equiv, 1h) and in the same reaction conditions as described above.

3.1.3.3.4 Thiol-specific Bis-Alkylation GST – reagent 4 mPEG-MS

For the PEGylation reaction using the bis-alkylation reagent **4,** the protein was first reduced in the same way as performed for PEG-MAL. After the removal of

DTT the reduced GST (0.5 mg/mL and 1mL) in 50 mM phosphate buffer pH 7.8 was then mixed with an increased molar excess of freshly prepared PEG-MS and the reaction was allowed to proceed for 1h at 7ºC with occasional shaking. Unreacted PEG was removed by cation exchange chromatography) as performed above for PEG-ALD reactions. PEGylation without reducing the protein was also performed to determine the extent of non-thiol specific PEGylation and if there were free protein thiols present. The purification steps were the same as described above. The GST-ZnF fusion protein was PEGylated using this reagent in similar conditions described for the GST (see Figure 6 for details).

3.1.3.4 Physico-chemical Characterization of the Conjugates

3.1.3.4.1 SDS-PAGE analysis

Purity of each PEGylation reactions was assessed by SDS-PAGE analysis using NuPAGE 4-12 % Bis-Tris Gel and MOPS buffer (Invitrogen, Carlsbad, CA, USA). Samples were prepared under reducing conditions and protein samples were resolved in 1 h at 200V. SDS-PAGE gels were stained with colloidal blue stain.

3.1.3.4.2 Size-exclusion HPLC

Each reaction was further characterized using size exclusion chromatography with a Hi-Load Superdex 200™ prep grade column (GE Healthcare). SEC was also used for the separation of protein-PEG conjugates from unreacted protein or other low molecular weight contaminants present in the sample. Typically 1.8 mL of sample solution was loaded into the column at a flow rate of 1 mL/min using 50 mM Tris HCl containing 150 mM NaCl, pH 7.4 as the eluent buffer. Protein components were detected at UV 280 nm.

3.1.3.4.3 MALDI-TOF-MS

Mass spectra were acquired using an Applied Biosystems Voyager System DE PRO MALDI-TOF mass spectrometer using nitrogen laser. The matrix was a saturated solution of α-cyano-4-hydroxycinnamic acid in a 50:50 mixture of acetonitrile and water containing 0.1% trifluoroacetic acid. The spectrum was

acquired in positive mode over the range 600-2500 Da using reflectron mode and 2-100 kDa using linear mode.

3.1.3.4.4 SDS-PAGE bands densitometry

PEGylation reaction efficiency (yield) and the percentage of each PEG-conjugate species obtained in each reaction were estimated by comparative bands densitometry analysis obtained from the colloidal blue stain stained SDS-PAGE gels using the ImageJ software [23; 24; 25]. Quantitative protein estimation by densitometry is not practical with silver stained gels since silver stain binds to protein non-stoichiometrically [26].

3.1.3.5 Affinity partitioning in aqueous two-phase systems

Aqueous two-phase systems composed by 6 % (w/w) PEG 10,000 – 8 % (w/w) DEX 500 were prepared by weighing appropriate amounts of stock solutions of the phase forming polymers in 2 mL graduated tubes. Varied native and PEGylated GST-ZnF protein concentration (see figures for details) were allowed to interact with the double-stranded DNA Oligo containing the ZnF binding site (6 μg / 3.34 ρmol) in 20 mM Tris-HCl buffer for at least 10 min at room temperature in 0.2 mL (final volume). The resulting complex was then added to the ATPS and pH was adjusted by adding 40 mL of 0.5 M Tris HCl buffer, pH 7.4. Finally, ultra pure water was added to make 1.5 g total system weight. The systems were then mixed by vortexing until all polymers were completely dissolved. Phase separation was accomplished by centrifugation (2000 × g, 2 min) and each phase was collected for further analysis. Phase turbidity was avoided by equilibrating at 30 ºC prior to collection of each phase sample. All further analyses were conducted at room temperature on the same day that the partitioning systems were prepared.

3.1.3.6 Protein Partition coefficient determination

The quantitative determination of the protein in each phase of these systems was performed by the Bradford assay using BSA as standard. The interference of PEG and DEX was overcomed by diluting the samples ten fold, and making a series of calibration curves using the diluted phase under analysis as solvent for

the BSA standards as described previously [27]. All concentrations were calculated using the average of at least 3 independent measurements. The partitioning of a protein in a two-phase system is described by the partition coefficient, K_c, and defined as $K_c = CT/CB$, where *C*T and *C*B are the concentrations of the protein in the top (PEG) and bottom (DEX) phases, respectively. Phase recovery (*PR*) of each protein, *x, in each phase (i)* was calculated as:

$$PR = \frac{[x]^i \times V_i}{x_{top} + x_{bottom}}$$

where V_i is the volume of the phase.

3.1.3.7 DNA partitioning analysis

Native PAGE electrophoresis analysis of the bottom and top phases of each system was performed to evaluate the double-stranded DNA partitioning and its binding to the modified GST-ZnF affinity protein. SDS-free Tris-glycine buffers were used and the electrophoresis performed as previously described [15].

3.1.4 Results and Discussion

The main goals of this study were to optimize four distinct PEGylation reactions on the GST-tag protein for later applications with a GST-fusion protein, to enhance the partitioning of a GST fusion ligand to the top phase of a PEG-dextran ATPS system and to ensure efficient PEGylation without adversely affecting binding activity to the target biomolecule. Four PEG reagents were examined, two of which are selective for protein amines and the other two for protein thiols.

3.1.4.1 Amine selective PEGylation of GST

Two different conjugation chemistries were used for the amine selective PEGylation of the GST protein, but both of these reagents are widely used in PEGylation: reagent **1** 20 kDa mPEG-ALD and **2** 20 kDa mPEG-SPA (Figure 1).

3.1.4.1.1 mPEG-ALD PEGylation of GST

The reductive alkylation reaction conditions using reagent **1** were optimized in terms of the concentration of the reactants, reaction time, the concentration of reducing agent $NaCNBH_3$ and reaction pH.

At pH 5.0, the extent of GST PEGylation was increased from ~5 up to 95 % (as determined by densitometry) by increasing the molar ratio of PEG-ALD to protein (PEG equiv) from 1 to 50, (Figure 2A), albeit with the formation of multi-PEGylated species. Mono-PEGylation was achieved using either 1 or 5 molar excess of reagent **1**, though the maximum PEGylation yield was 12 % at this stoichiometry. Reaction with 50 PEG equiv resulted in a yield above 90 % with PEGylation at multiple sites of conjugation (see SDS-PAGE inset in Figure 2A).

The reaction time used for PEG conjugation is also an important parameter (Table 1). PEGylation yield for the 25 equiv excess of PEG-ALD reactions, increased from 74 % to 98 % in the period of 7 to 72h but was accompanied by greater product heterogeneity. This reagent is more reactive at pH 7.8 compared to pH 5 (Figure 2B) and though the yield increased from 32 to 42 % using 5 PEG-ALD equiv/24h reactions, at the higher pH di-PEGylated GST appeared.

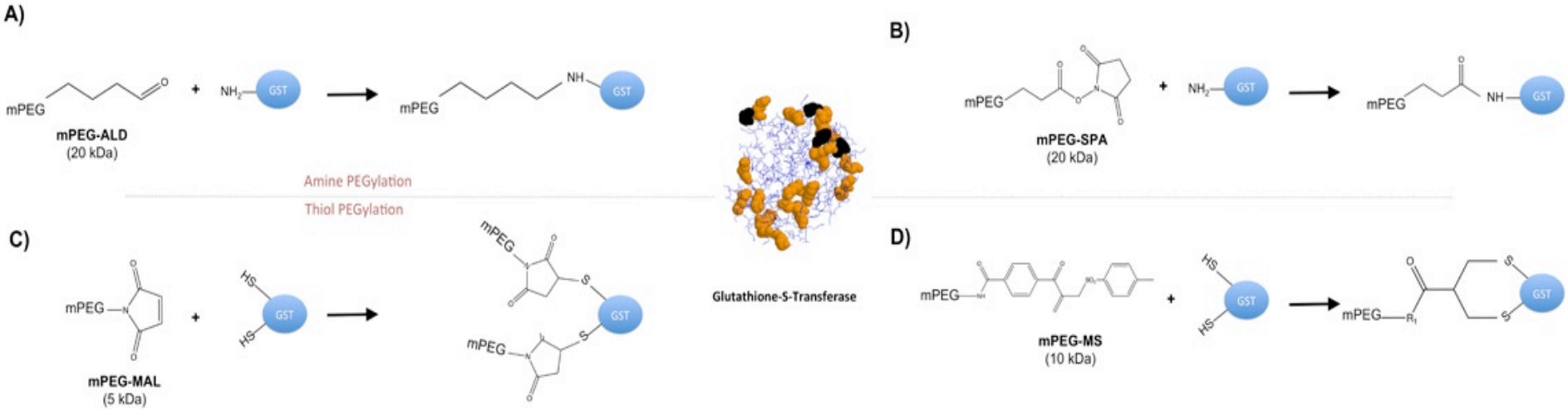


Figure 1 – GST PEGylation reaction schemes for the amine (A and B) and cysteine thiol (C and D) PEG conjugation. A) 1 mPEG-ALD reaction; B) 2 mPEG-SPA reaction; C) 3 mPEG-MAL reaction; D) 4 mPEG-MS reaction. In the centre is represented the 3D structure of GST protein (Protein data bank code: 1GTA). The four cysteines amino acids within GST are highlighted as black balls and the 21 lysines amino acids as orange balls.

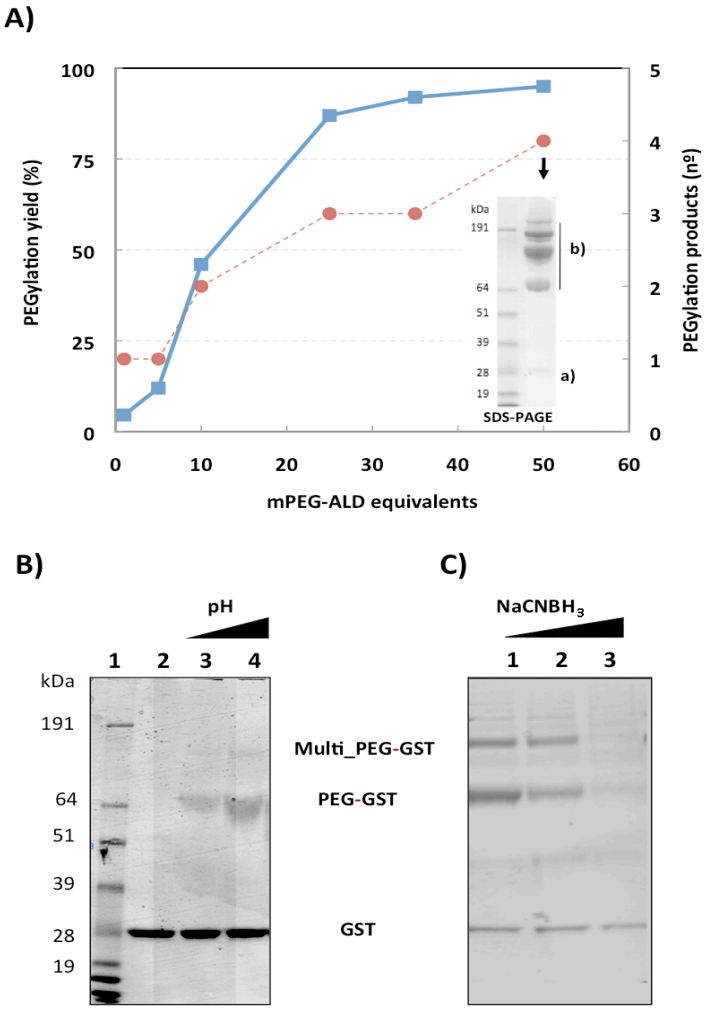


Figure 2 - Characterisation of GST PEGylation using reagent 1, mPEG-ALD. A) GST PEGylation efficiency (square) and number of the PEGylation products (circle) with increased mPEG-ALD equivalents (3 mg/mL $NaCNBH_3$, pH 5, 24h reaction) Inset: SDS-PAGE reaction characterization using 50 equivalents of reagent 1 showing a) free GST and b) GST PEGylated products. B) SDS-PAGE gel of the PEGylation reaction at different pH using 5 PEG equivalents over 24h reaction: 1) Molecular weight markers; 2) free GST; 3) PEGylation at pH 5; 4) PEGylation at pH 7.8. C) SDS-PAGE gel showing PEGylation reactions with increased $NaCNBH_3$ concentration: 1) 1mg/mL; 2) 3 mg/mL; 3) 6mg/mL.

Reagent **1** is known to react specifically with the N-terminal residue of some proteins in acidic pH (~5) because of the potential difference between the pKa of the terminal amine- (7.6 – 8.0) and the ε-amines (10.0 – 10.2) [3]. Conjugation at the protein terminal residue would likely minimize any interference between PEG and important amino acid residues for biological activity but, as seen here, mono-PEGylation is commonly achieved at low reaction yield.

Table 1 - GST mPEG-ALD PEGylation yields with the increase of the reaction time using 25 mPEG-ALD equiv in 50 mM acetate buffer pH 5 and 1 mg/mL $NaCNBH_3$.

Reaction time (h)	PEGylation products (nº)	PEGylation yields (%)
0	0	0
7	2	74.6
24	3	85.4
48	3	92.0
72	4	98.3

Barbosa *et al*. [15]have shown that the multi-PEGylation of a DNA binding protein did not disrupt DNA binding and similar binding constants were obtained compared to the unconjugated protein. Moreover, the authors have shown that the utilization of heterogeneous PEGylated species could be efficiently used in ATPS affinity applications [15]. This represents an advantage over the utilization of mono-PEGylated ligands since the separation of different PEGylated species is a complicated process and the cost of activated PEGs for efficient mono-PEGylation reactions is high [28]. Hence, the focus here is on the optimization of PEGylation yield rather than achieve homogenous PEGylation products.

The effect of sodium cyanoborohydride ($NaCNBH_3$) concentration as reducing agent is shown in Figure 2C. A decrease in PEGylation yield was observed with increased concentration of reducing agent (from 1 to 6 mg/mL). The best reaction yields were obtained when 1mg/mL concentration (15.9 mM) was used per 0.4 mg/mL GST (13.9 μM). Excess reducing agent might make the solution alkaline, promoting GST aggregation.

3.1.4.1.2 mPEG-SPA PEGylation of GST

Figure 3A shows a SDS-PAGE analysis of reaction products obtained with increased PEG-SPA/GST stoichiometry at pH 7.8/ 24h. Compared to the reductive amination strategy with reagent **1,** these reactions with the acylation reagent **2** were less efficient giving lower yields using similar PEG equiv. When

a 100 molar excess of reagent **2** was used only ca. 53 % of the protein was conjugated resulting in a population distribution of native (~47 %), mono- (~48 %) and di-PEGylated GST (~4 %).

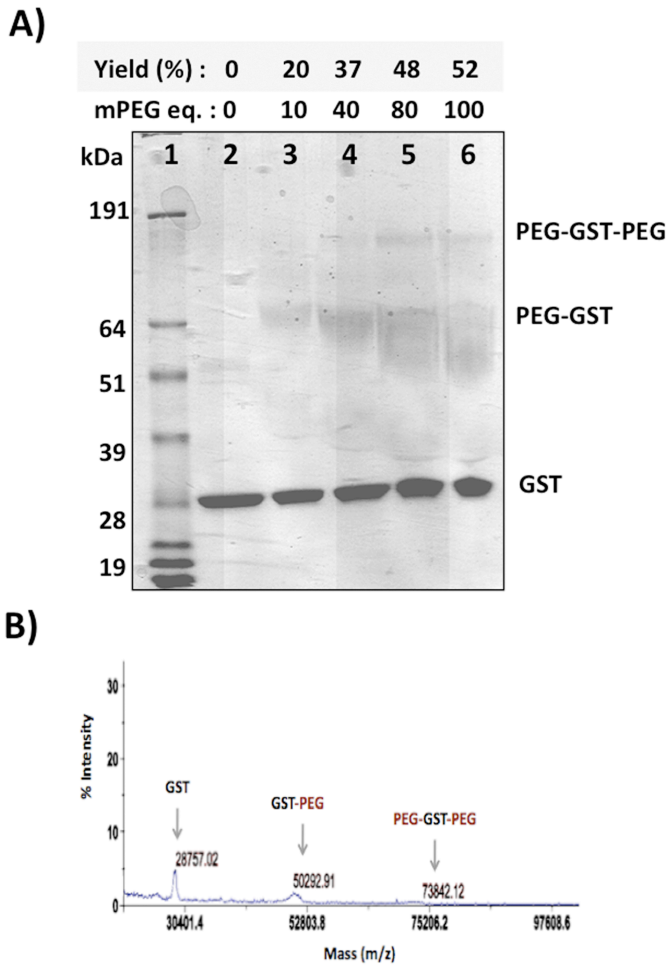


Figure 3 - Characterisation of the GST PEGylation reactions using reagent 2, mPEG-SPA. A) SDS-PAGE gel showing PEGylation reactions with increase in mPEG-SPA equivalents. B) MALDI-TOF mass spectrum of a typical 20 kDa mPEG-SPA reaction showing the presence of unreacted GST, mono-PEGylated GST and di-PEGylated GST.

The reaction yield increased as the equiv of reactive PEG increased but mono-PEGylation was only achieved with only 20 % yield. Since this reagent has a short hydrolytic half-life (16.5 min) in aqueous solutions no reactions were conducted with longer reaction times. Instead, reagent **2** was added as a powder directly to the stirred protein solution, but no significant improvements were observed compared to aqueous reaction (data not shown). MALDI-TOF analysis of a typical 20kDa PEG-SPA reaction (Figure 3B shows native GST (28.7 kDa) together with mono- and the di-PEGylated GST species with the

proportional increment of PEG-SPA mass (20,000 Da) for the sequential series of peaks in the spectrum. This result is in agreement with the SDS-PAGE gel.

The difference in reactivity between the two activated PEGs **1** and **2** has been observed for other proteins [3; 29] and the distinction is due to the different synthetic routes for conjugation: alkylation (**1**) vs. acylation (**2**). The reductive alkylation conjugation occurs via formation of Schiff's base, which requires reduction *in situ* with sodium cyanoborohydride to yield a stable amine linkage. PEGylation by acylation removes the charge of the α- amino group through the formation of an amide bond. In addition, whereas imine formation in water is favoured by the addition of increased amounts of reagent **1**, competitive hydrolysis of reagent **2** removes this reagent from the reaction pathway.

3.1.4.2 Thiol selective PEGylation of GST

Two reagents were evaluated for thiol-specific PEGylation of the four cysteines on GST: Reagent **3**, a 5 kDa mPEG-MAL, and **4** 10 kDa mPEG-DB (Figure 1). Initial investigation focused on the specific reactivity of these two reagents without reducing the disulfides in the protein. Comparative reactions were then conducted using protein that was reduced using 100 mM DTT immediately prior to the PEG conjugation reactions. Prior to PEGylation the DTT was removed by buffer-exchange with PD-10 columns.

3.1.4.2.1 mPEG-MAL PEGylation of GST

SDS-PAGE characterisation of PEG-conjugation to native and reduced GST using reagent **3** is shown in Figure 4. These reactions achieved completion faster (~1 h) than those with the amino reactive PEGs (>24 h) under reducing conditions (lane 4, Figure 4a) compared with 85 % PEGylation for the non-reducing conditions (lane 2). However, non-reducing conditions resulted in four different PEGylation products. This may correspond to proteins with a different number of PEGs attached to the four Cys amino acids residues.

Nonetheless, the protein band corresponding to tetra-PEGylated GST represents only 3 % of all reacted GST, whereas the di-PEGylated GST predominates with conversions of 72 % (c.f. mono-PEGylation 13 % and tri-PEGylated 12), suggesting that only two free thiols are easily accessible for PEGylation.

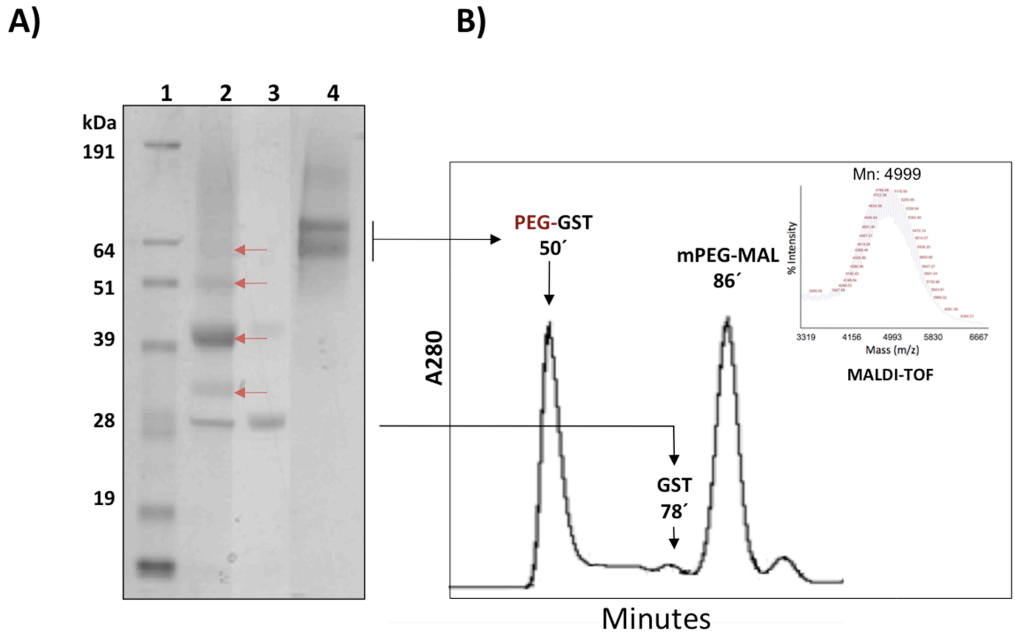


Figure 4 – Characterisation and isolation of GST protein PEGylated with reagent 3, mPEG-MAL. A) SDS-PAGE showing the PEGylation of non-reduced GST with 20 PEG equivalents of mPEG-MAL (lane2) pH 7.8 (1h), free GST protein (lane 3) and 20 equivalents mPEG-MAL using reduced GST pH 7.8 (1h). B) Size exclusion chromatogram of the GST PEGylation reaction shown in lane 4, showing the removal of excess of mPEG-MAL reagent (86´ peak) and non-reacted GST (~78´). Inset shows the MALDI-TOF analysis of the 86´ SEC peak showing an average Mn of 4,999 Da corresponding to the non-reacted mPEG-MAL.

As for other PEGylated GST proteins, the SDS-PAGE gel shows that the PEG conjugated protein migrated at much higher apparent MW than expected by comparison with the protein marker. This feature of PEGylated proteins has been observed before by Kurfurst [30]. Although, the PEGylated proteins showed linear mobility with increase MW during migration on SDS PAGE gels, the comparison of the apparent molecular weights of conjugated species and the non-PEGylated protein standards resulted in incorrect molecular weight estimation due to the decrease of the electrophoresis mobility of PEGylation proteins.

The size exclusion chromatogram (SEC) of the reaction products obtained with 20 molar excess of PEG per reduced GST (5 molar excess per Cys) is shown in Figure 4B. When using low MW activated PEGs such as the 5 kDa reagent **3**, a single chromatography step can be used to remove both the non-reactive GST and the excess of PEG using a size exclusion column. For higher MW PEGs (> 10 kDa) their retention times are lower than that of GST and cation exchange chromatography has to be used to remove unreacted PEGs (data not shown). To confirm that the peak eluted after GST was the unconjugated PEG species MALDI-TOF analysis of this fraction was obtained which showed a mass at 5 kDa corresponding to reagent **3** (Figure 4B inset).

3.1.4.2.2 mPEG-MS PEGylation of GST

The results of reactions using reagent **4** are shown in Figure 5. As with the maleimide PEGylation it was only observed almost complete conversions when the protein was fully reduced Figure 5 (A and B). Using 10 PEG equiv the reaction efficiency was around 34 % at pH 7.8 for non-reduced GST, whereas the reaction efficiency was above 90 % when reducing conditions were used.

Again, only two Cys moieties were accessible for the PEGylation when non-reduced conditions were used since only PEGylated protein product was observed. In this case the reaction yields mainly mono-PEGylated protein as this reagent is capable of forming a 3-carbon bridge between two sulfhydryls. Similarly to the other PEG reagents, as the PEG molar ratio was increased the number of different PEGylated species also increased. Mono-PEGylation resulted using 1PEG equiv but heterogeneous population of PEGylated species were obtained when 15 PEG equiv were used (Figure 5A, lane 5).

The separation of the PEGylated conjugates using either 1 or 15 PEG equiv of reagent **4** was performed by size exclusion chromatography as shown in Figure 5C. Peaks eluted with different retention times dependent upon the size of the PEG-conjugate, as shown. Similar to the SDS-PAGE gel in Figure 5A it was seen that the peak at 58 min, corresponding to the di- and tri-PEGylated species,

increased with the increase of PEG equiv used (Figure 5C-2). Non-reacted GST (78 min) was separated from the conjugates and the mono-PEGylation species (64 min), but the HPLC peak resolution inadequate to separate the other PEG-conjugates.

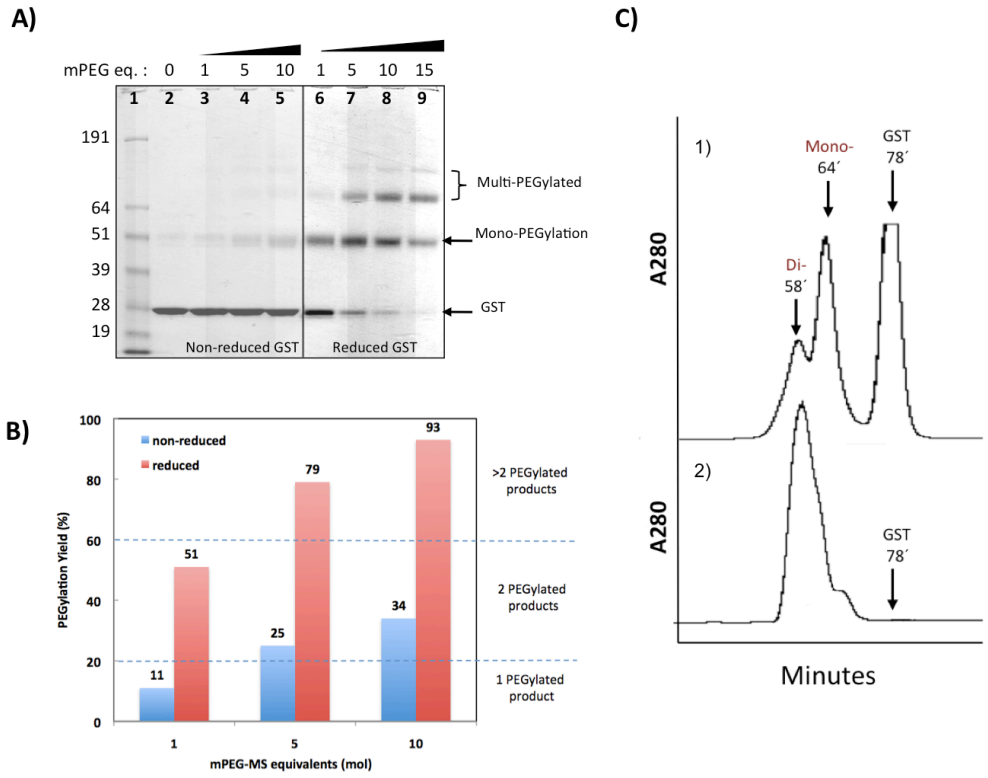


Figure 5 – Characterization and isolation of GST protein PEGylated with reagent 4, mPEG-MS A) SDS-PAGE Gel showing the mPEG-MS reaction with increased molar ratio of mPEG-MS using reduced and non-reduced protein. B) Reaction efficiency under reduced and non-reduced conditions with increased PEG equivalents. The number of PEGylation products is shown for each reaction. C) Size-exclusion chromatograms of the PEGylation reactions with reduced GST using 1 PEG equivalent (1) or 15 PEG equivalents (2) in 50 mM phosphate buffer pH 7.8, 1h reaction time.

3.1.4.3 PEGylation of GST-ZnF fusion protein

The most efficient PEG conjugation conditions for GST were applied to GST-ZnF (MW 38.3 kDa, [22]). The zinc finger protein increases the total protein size by 99 amino acids. The zinc-finger has an additional 6 cysteines and 13 Lys for PEGylation. Thus, the number of potential PEGylation sites increased from 21 to 34 Lys (for amine PEGylation) and from 4 to 10 Cys (for thiol PEGylation) from GST to GST-ZnF. In consequence, it was anticipated that more complex mixtures of PEGylated products might result with the GST-ZnF fusion protein. A

further complication was that the free GST protein was co-expressed with GST-ZnF although at much lower concentrations (see lower band around 28 kDa, Figure 6A). This has been observed previously [14] and attempts to remove the free GST protein in solution by either SEC or Ionic-exchange chromatography proved infeasible. Thus, GST had to be taken into account in considering the product heterogeneity.

SDS-PAGE gels for PEG-conjugation reactions of GST-ZnF using reagents **1, 3** and **4** are shown in Figure 6A. Reagent **2** mPEG-SPA was not evaluated in view of its low conjugation efficiency compared to the other activated PEGs.

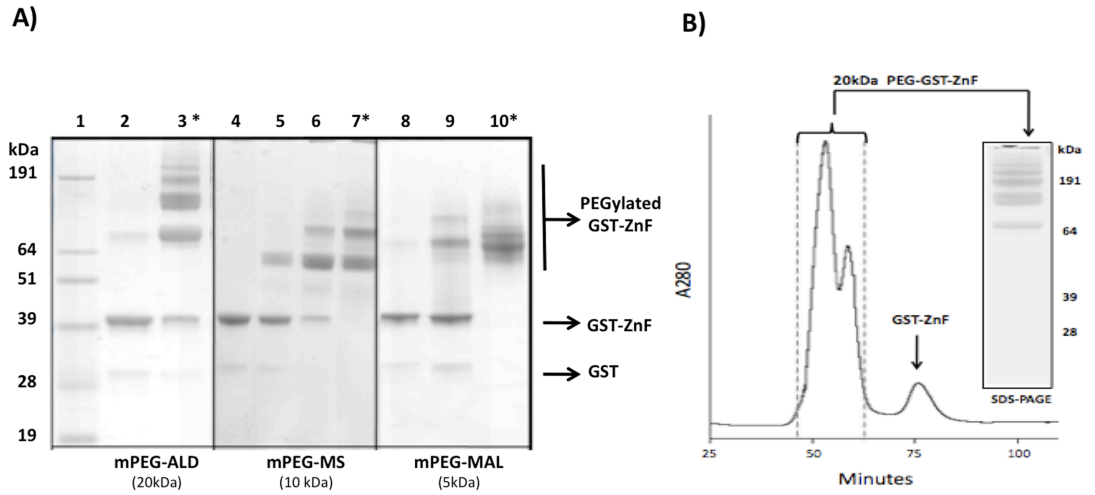


Figure 6 - Characterisation of the GST-ZnF fusion protein PEGylation with the mPEG-ALD, mPEG-MS, and mPEG-MAL reagents. A) SDS-PAGE gels showing the reaction products for each reagent: 1) Marker; 2) 1 mPEG-ALD equiv (pH, 24h); 3) 25 mPEG-ALD equiv (pH 5, 48h); 4) free GST; 5) 1 mPEG-MS equiv (a); 6) 5 mPEG-MS equiv (a); 7) 10 mPEG-MS equiv (a); 8) free GST; 9) non-reduced GST- 20 mPEG-MAL equiv (a); 10) reduced GST- 20 mPEG-MAL equiv. (a). B) Size exclusion chromatogram for the isolation of the GST-ZnF PEGylated products of lane 3 Fig. 6A reaction. Inset shows the SDS-PAGE gel with the collected peak limited within the dashed lines. * Reaction products from these reactions were used for the partitioning studies in ATPS. (a) Reactions were performed at pH 7.8 for 1h.

The PEG-conjugation reactions with GST using **1** resulted in yields above 90 % when reaction time was increased to 48h (using 5 equiv) or by using 25 PEG equiv in 24 h at pH 5. For the GST-ZnF PEGylation, 25 PEG-ALD equiv were used and the reaction was allowed to proceed over a 48h reaction period at pH 5 at a temperature of 7ºC. The reaction yield using these conditions was about 80 % and resulted in multi-PEGylated species as shown in Figure 6A, lane 3.

The reaction conditions selected here involved a compromise between achieving a relatively low product heterogeneity and high reaction yield. It has been demonstrated already that almost complete conjugation reactions (95.3%) after 3 days of reaction time using this same protein and the reagent **1** [15]. However, more than 8 products of reaction were obtained.

A typical SEC chromatogram for the separation of unreacted GST-ZnF from the reaction using 25 PEG equiv of reagent **1** mPEG-ALD is shown in Figure 6B. Peak fractions were collected and analysed using SDS-PAGE (Figure 6b inset). As it proved difficult to resolve the different PEGylated species, fractions were collected as shown and combined for further analysis.

Thiol specific PEGylation of the 10 Cys sites on GST-ZnF was also evaluated using reagents **3** and **4.** About half of the fusion protein (47 %) underwent reaction with reagent **3** PEG-MAL when mixed directly with the protein in non-reducing conditions (lane 9, Figure 6). Almost ~100 % of the protein underwent reaction in 1h when it was first reduced with DTT and then allowed to react with a 20 molar excess of PEG-MAL **3** at pH 7.8 (lane 10, Figure 6). The PEGylated products using these conditions were separated and used for partitioning studies in PEG/dextran systems.

Disulfide binding reagent **4** also underwent similar reactions with GST-ZnF. Using 1 molar equiv the mono-PEGylation could be achieved using reduced conditions (yield 39 %), but complete reactions were achieved when using only 10 molar excess of PEG resulting in multi-PEG conjugates (lane 7, Figure 6). These later reactions were chosen to perform the following studies after purification by ion exchange and SEC.

.

3.1.4.4 Protein partition in ATPS

An ATPS system comprised of PEG 10,000 and DEX 500 was selected to evaluate the GST-ZnF protein partition before and after PEGylation. Before PEGylation the fusion protein partitioned more than 90 % (k= 0.036) to the

bottom phase whereas after PEGylation it partitioned predominantly to the top phase of these same systems (Figure 7A and B). GST-ZnF PEGylated using 20 kDa reagent **1** partitioned 96 % to the top phase (K= 8.1); 97.1 % for the 5 kDa reagent **3** (K=11.0); and 97.2 % for the 10 kDa reagent **4** (K=11.6).

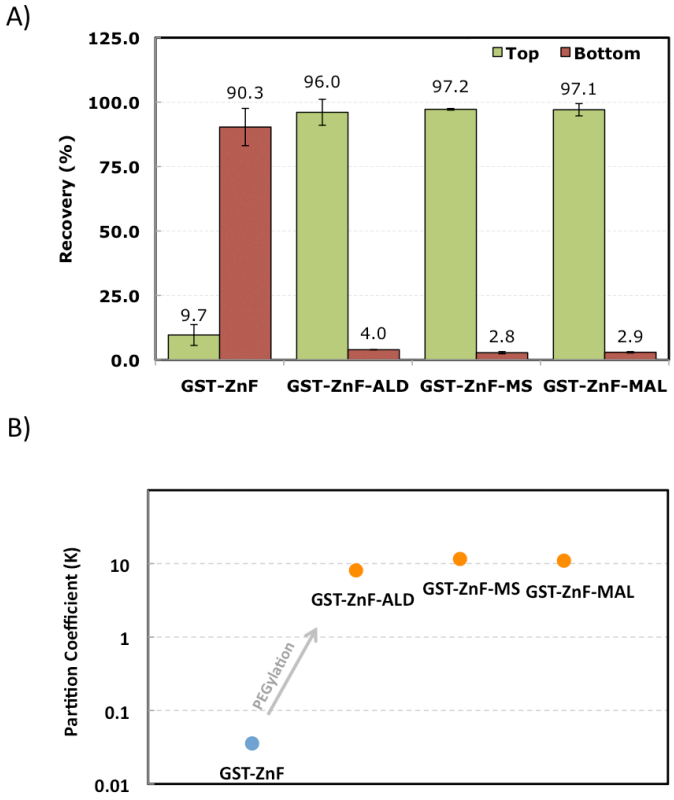


Figure 7 - Modified and non-modified GST partitioning in PEG 10,000 – DEX 500 pH 7.4. A) Protein phase recovery (%). B) Protein partition coefficient ($K_{protein}$) in each system.

In a previous study the partitioning of GST-ZnF PEGylated with reagent **1** (using less PEG equiv, 5) was examined in a PEG 600 – DEX 40 ATPS and similar results to those presented here were obtained as the PEG-GST-ZnF partitioned 97.5 % to the top phase [15].

Effective affinity partitioning using PEGylated ligands requires that the affinity ligand steers the partitioning of the target biomolecule to the PEG phase as shown in Figure 8. It is shown here the ligand accumulates in top phase of these ATPS in minimal loss to the bottom phase (less 3 %) after PEGylation and any of the activated PEGs used can be successfully used to this end. However, the highly heterogeneous population of PEG-GST-ZnF species makes it difficult to

compare the partitioning of the PEGylated proteins in terms of the effect of the mPEG molecular weight or reaction chemistry. Nevertheless, a linear relationship between the partition coefficient in ATPS and the degree of PEGylation has also been reported [31].

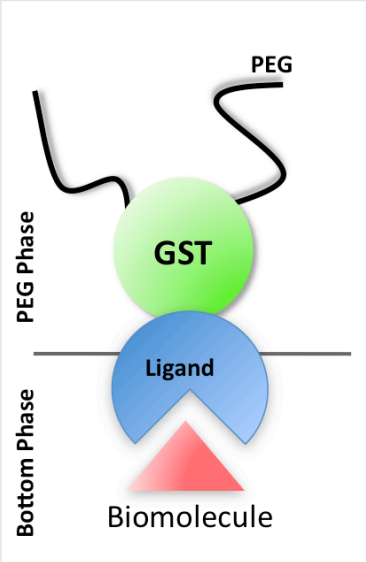


Figure 8 - Representation of the PEGylated GST-fused ligand for biomolecule affinity separation in aqueous two-phase systems.

3.1.4.5 DNA affinity partition in ATPS

While the ligand partitioning to the top phase is fundamental, the conjugated PEG might inhibit the binding activity of the ligand and thus make it ineffective for affinity extractions in ATPS. The affinity partitioning of the 29 base pairs (bp) double-stranded DNA oligonucleotide containing the zinc-finger binding site was evaluated using the PEGylated and non-PEGylated GST-ZnF protein ligand in aqueous two-phase systems composed by PEG 10,000 and DEX 500. Evaluation of the effect of the differently targeted PEG-conjugations on the DNA binding activity of the GST-ZnF protein and the top and bottom phase composition of these systems was conducted by native PAGE gels stained for DNA (Figure 9). When DNA binds to the GST-ZnF protein a shift of the free DNA band to a higher molecular weight is observed. A yet higher shift occurs (lower electrophoretic mobility) when PEGylated protein is used since the PEG increases the total mass of the binding complex.

As seen in Figure 9A the double-stranded DNA oligonucleotide partitioned to both phases when no protein was present in the system. When GST-ZnF was added the DNA bound to the ligand and partitioned to the bottom phase since the protein accumulates preferentially in this phase (Figure 9B). On the other hand, it was expected that a DNA band shift would occur with samples of the top phase when PEGylated GST-ZnF was used since the PEGylated ligand accumulates in this phase. As seen in Figure 9 (2, 3 and 4), only systems with PEGylated GST-ZnF made using reagent **1** (amine PEGylation) resulted in a DNA band shift in samples from the top phase and no evident interaction in either top or bottom phases were observed when the fusion protein was PEGylated using the thiol-specific reagents **3** and **4**.

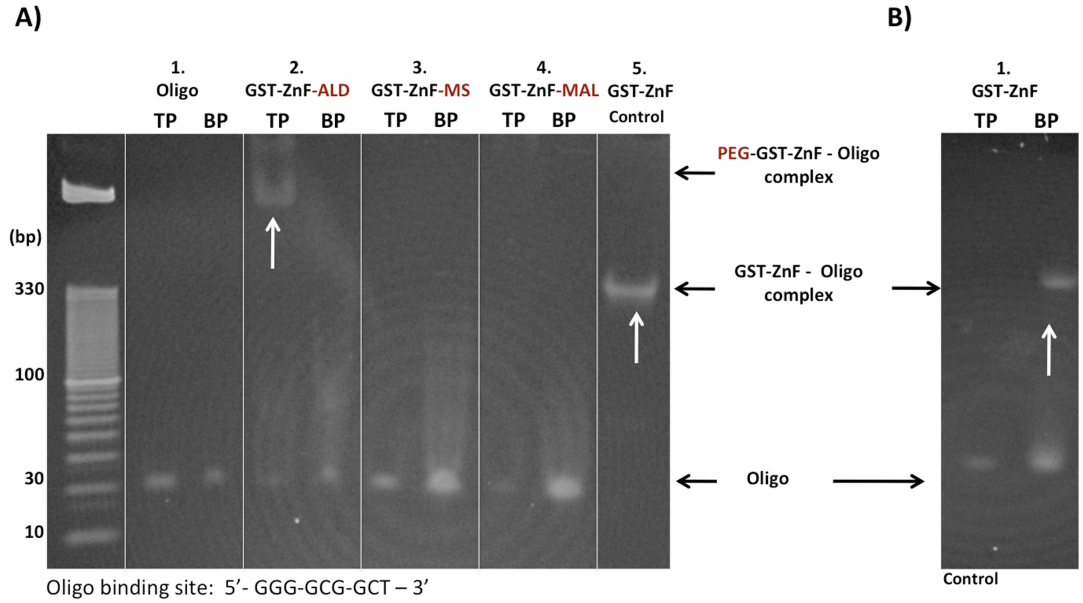


Figure 9 - Native PAGE electrophoresis of the top (TP) and bottom (BP) phases of PEG 10,000 – DEX 500 systems composed of double stranded DNA oligonucleotide bearing binding site for GST-ZnF protein. A) Analysis of systems composed of different PEGylated GST-ZnF protein and the DNA oligonucleotide. Arrows denotes DNA/protein binding complex in the gel. B) Analysis of system composed of unmodified GST-ZnF fusion protein and the DNA oligonucleotide. Gels stained with ethidium bromide for DNA.

In each aqueous two-phase system it was used 2.24 μM of the modified protein per 3.34 pmol of DNA, which is 2-fold molar protein excess as used on the positive control shown in lane 5, Figure 9. The control was prepared by mixing 1.12 μM of unmodified GST-ZnF and 3.34 pmol of DNA oligonucleotide in

homogenous solution, without two-phase system. In the ATPS where thiol PEGylation proteins were used an even higher protein/DNA ratio was tested (5-fold more) but no binding interaction was observed.

The results from the present study for thiol-specific PEGylation are in agreement with earlier reported work, which described the DNA binding inhibition in buffer solutions after the fusion GST-ZnF has been PEGylated using reagent **3**.

Nonetheless, it has been shown that both reagents that undergo alkylation (**3**) or bis-alkylation (**4**) had been successfully applied for the PEG conjugation of therapeutic proteins, enzymes or antibody fragments [1; 18; 32; 33]. PEGylation did not disrupt either their tertiary structure or their biological activity of these proteins, although none have Cys amino acids on the active site.

3.1.5 Conclusions

The results demonstrate that each of the three PEGylation reactions performed with the GST-ZnF fusion protein resulted in similar PEGylation yields to that obtained for the free GST-tag protein. Both thiol selective PEG reactions were completed after 1h under reducing conditions using either GST or GST-ZnF fusion protein. However, reaction yields dropped below 50 % when non-reducing conditions were used. By contrast, amine selective reagents required much longer reaction times (24h) or high PEG equiv, in order to yield more then 80 % conjugation, which could be prohibitive for large-scale affinity applications.

All three PEGylated GST- ZnF fusion proteins partition almost completely (>96 %) to the top phase of PEG-based systems as opposed to non-PEGylated (only 10 %), which is desirable for the increased selectivity of an affinity extraction process in ATPS. Nevertheless, it is demonstrated that the selection of a suitable PEG regent has to be performed with caution to avoid interference of the conjugated PEG with the ligand-binding site. Both thiol PEGylation methods disrupted the GST-ZnF binding to the DNA bearing the ZnF recognition site due to PEGylation on the cysteine residues at the ZnF DNA binding site. However, amine PEGylation of GST-ZnF protein did not disrupt the DNA binding activity of the ZnF moiety resulting in the affinity partitioning of the DNA from the bottom to the top phase of these systems. Since most binding sites do not have free thiols, it is anticipated that the thiol reactive PEG reagents would be more appropriate for other affinity ligands. Amine selective reagents would be suitable for ligands with Cys at the ligand-binding site.

It is hoped that our findings will aid in the development of affinity purification processes using PEGylated GST fusion ligands in ATPS and thus, improving the low selective of these systems.

3.1.6 References

[1]S. Brocchini, A. Godwin, S. Balan, J.W. Choi, M. Zloh, and S. Shaunak, Disulfide bridge based PEGylation of proteins. Adv Drug Deliv Rev 60 (2008) 3-12.

[2]P.I. Guerra, C. Acklin, A.A. Kosky, J.M. Davis, M.J. Treuheit, and D.N. Brems, PEGylation prevents the N-terminal degradation of megakaryocyte growth and development factor. Pharm Res 15 (1998) 1822-7.

[3]H. Lee, I.H. Jang, S.H. Ryu, and T.G. Park, N-terminal site-specific mono-PEGylation of epidermal growth factor. Pharm Res 20 (2003) 818-25.

[4]J. Medina, L. Garcia Buey, J.A. Moreno Monteagudo, M. Trapero Marugan, and R. Moreno Otero, Therapeutic advantages of pegylation of interferon alpha in chronic hepatitis C. Rev Esp Enferm Dig 95 (2003) 568-74, 561-7.

[5]H.-K. Rajni, Aqueous Two-Phase Systems: Methods and Protocols (Methods in Biotechnology), Humana Press, Inc, Totowa, New Jersey, 2000.

[6]L.J. Karr, J.M. Van Alstine, R.S. Snyder, S.G. Shafer, and J.M. Harris, Cell separation by immunoaffinity partitioning with polyethylene glycol-modified protein A in aqueous polymer two-phase systems. J Chromatogr 442 (1988) 219-27.

[7]L.J. Karr, D.L. Donnelly, A. Kozlowski, and J.M. Harris, Use of poly(ethylene glycol)-modified antibody in cell extraction. Methods Enzymol. 228 (1984) 377-90.

[8]A.F. Buckmanm, M. Morr, and M.R. Kula, Preparation of technical grade polyethylene glycol (Mr 20.000)-N6-(2-aminoethyl)-NADH by a procedure adaptable to large-scale synthesis. Biotechnol. Appl. Biochem. 9 (1987) 258-268.

[9]L.O. Person, and B. Olde, Synthesis of ATP-polyethylene glycol and ATP-dextran and their use in the purification of phosphoglycerate kinase from spinach chloroplasts using affinity partitioning. J. Chromatogr. 457 (1988) 183-193.

[10]G. Birkenmeier, V.P. Shanbhag, and G. Kopperschläger, Comparison of the binding of poly(ethylene glycol)-bound fatty acids to human albumin and alpha-fetoprotein studied by affinity phase partitioning. Biomed Biochim Acta. 45 (1986) 285-9.

[11]G. Takerkart, E. Segard, and M. Monsigny, Partition of trypsin in two-phase systems containing a diamidino-alpha,omega-diphenylcarbamyl poly (ethylene glycol) as competitive inhibitor of trypsin. FEBS Lett. (1974) 218-220.

[12]M. Carey, and S.T. Smale, Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques, Cold Spring Harbor Laboratory Press,, Cold Spring Harbor, NY, 2000.

[13]J.V. Frangioni, and B.G. Neel, Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins. Anal Biochem 210 (1993) 179-87.

[14]S. Ghose, G.M. Forde, and N.K. Slater, Affinity adsorption of plasmid DNA. Biotechnol Prog 20 (2004) 841-50.

[15]H. Barbosa, A.V. Hine, S. Brocchini, N.K. Slater, and J.C. Marcos, Affinity partitioning of plasmid DNA with a zinc finger protein. J Chromatogr A O 1206 (2008) 105-12.

[16]O. Kinstler, G. Molineux, M. Treuheit, D. Ladd, and C. Gegg, Mono-N-terminal poly(ethylene glycol)-protein conjugates. Adv Drug Deliv Rev 54 (2002) 477-85.

[17]M.J. Roberts, M.D. Bentley, and J.M. Harris, Chemistry for peptide and protein PEGylation. Adv Drug Deliv Rev 54 (2002) 459-76.

[18]S. Balan, J.W. Choi, A. Godwin, I. Teo, C.M. Laborde, S. Heidelberger, M. Zloh, S. Shaunak, and S. Brocchini, Site-specific PEGylation of protein disulfide bonds using a three-carbon bridge. Bioconjug Chem 18 (2007) 61-76.

[19]S. Shaunak, A. Godwin, J.W. Choi, S. Balan, E. Pedone, D. Vijayarangam, S. Heidelberger, I. Teo, M. Zloh, and S. Brocchini, Site-specific PEGylation of native disulfide bonds in therapeutic proteins. Nat Chem Biol 2 (2006) 312-3.

[20]J.R. Desjarlais, and J.M. Berg, Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. Proc Natl Acad Sci U S A 90 (1993) 2256-60.

[21]S. Brocchini, S. Balan, A. Godwin, J.W. Choi, M. Zloh, and S. Shaunak, PEGylation of native disulfide bonds in proteins. Nat Protoc 1 (2006) 2241-52.

[22]J. Woodgate, D. Palfrey, D.A. Nagel, A.V. Hine, and N.K. Slater, Protein-mediated isolation of plasmid DNA by a zinc finger-glutathione S-transferase affinity linker. Biotechnol Bioeng 79 (2002) 450-6.

[23]W.S. Rasband, ImageJ, US National Institute of Health, Bethesda, MD, 1997.

[24]A. Deiters, T.A. Cropp, D. Summerer, M. Mukherji, and P.G. Schultz, Site-specific PEGylation of proteins containing unnatural amino acids. Bioorg Med Chem Lett 14 (2004) 5743-5.

[25]J. Ramon, V. Saez, R. Baez, R. Aldana, and E. Hardy, PEGylated interferon-alpha2b: a branched 40K polyethylene glycol derivative. Pharm Res 22 (2005) 1374-86.

[26]G. Walsh, Pharmaceutical Biotechnology - Concepts and Applications, John Wiley & Sons, Ltd, West Sussex PO19 8SQ, England 2007.

[27]H. Barbosa, N.K. Slater, and J.C. Marcos, Protein quantification in the presence of poly(ethylene glycol) and dextran using the Bradford method. Analytical Biochemistry (2009).

[28]C.J. Fee, and J.M. Van-Alstine, PEG-proteins: Reaction engineering and separation issues. Chemical Engineering Science 61 (2006) 924-939.

[29]O.B. Kinstler, D.N. Brems, S.L. Lauren, A.G. Paige, J.B. Hamburger, and M.J. Treuheit, Characterization and stability of N-terminally PEGylated rhG-CSF. Pharm Res 13 (1996) 996-1002.

[30]M.M. Kurfurst, Detection and molecular weight determination of polyethylene glycol-modified hirudin by staining after sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Anal Biochem 200 (1992) 244-8.

[31]C. Delgado, M. Malmsten, and J.M. Van Alstine, Analytical partitioning of poly(ethylene glycol)-modified proteins. J Chromatogr B Biomed Sci Appl 692 (1997) 263-72.

[32]A.M. Cantin, D.E. Woods, D. Cloutier, E.K. Dufour, and R. Leduc, Polyethylene glycol conjugation at Cys232 prolongs the half-life of alpha1 proteinase inhibitor. Am J Respir Cell Mol Biol 27 (2002) 659-65.

[33]A.P. Chapman, PEGylated antibodies and antibody fragments for improved therapy: a review. Adv Drug Deliv Rev 54 (2002) 531-45.

Subchapter 3.2

Affinity Partitioning of Plasmid DNA with a Zinc Finger Protein

Journal of Chromatography A, 1206 (2008) 105–112

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Affinity partitioning of plasmid DNA with a zinc finger protein

H. Barbosa [a], A.V. Hine [b], S. Brocchini [c], N.K.H. Slater [d], J.C. Marcos [a,*]

[a] Department of Chemistry, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
[b] School of Biological Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK
[c] Department of Pharmaceutics, The School of Pharmacy, University of London, 29/39 Brunswick Square, London WC1N 1AX, UK
[d] Department of Chemical Engineering, University of Cambridge, Pembroke Street, Cambridge CB2 3RA, UK

ARTICLE INFO

Article history:
Received 18 February 2008
Received in revised form 22 July 2008
Accepted 23 July 2008
Available online 12 August 2008

Keywords:
Aqueous two-phase systems
PEGylation
Protein–DNA interactions
Zinc finger
Plasmid purification

ABSTRACT

The affinity isolation of pre-purified plasmid DNA (pDNA) from model buffer solutions using native and poly(ethylene glycol) (PEG) derivatized zinc finger–GST (Glutathione-*S*-Transferase) fusion protein was examined in PEG–dextran (DEX) aqueous two-phase systems (ATPSs). In the absence of pDNA, partitioning of unbound PEGylated fusion protein into the PEG-rich phase was confirmed with 97.5% of the PEGylated fusion protein being detected in the PEG phase of a PEG 600–DEX 40 ATPS. This represents a 1322-fold increase in the protein partition coefficient in comparison to the non-PEGylated protein (K_c = 0.013). In the presence of pDNA containing a specific oligonucleotide recognition sequence, the zinc finger moiety of the PEGylated fusion protein bound to the plasmid and steered the complex to the PEG-rich phase. An increase in the proportion of pDNA that partitioned to the PEG-rich phase was observed as the concentration of PEGylated fusion protein was increased. Partitioning of the bound complex occurred to such an extent that no DNA was detected by the picogreen assay in the dextran phase. It was also possible to partition pDNA using a non-PEGylated (native) zinc finger–GST fusion protein in a PEG 1000–DEX 500 ATPS. In this case the native ligand accumulated mainly in the PEG phase. These results indicate good prospects for the design of new plasmid DNA purification methods using fusion proteins as affinity ligands.

1. Introduction

Molecular therapies and vaccines based on purified plasmid DNA (pDNA) stimulate the need to develop scalable and cost-effective purification processes that can meet the stringent specifications on purity and quality [1,2]. Several chromatography-based processes have been proposed [3–5], as have non-chromatography processes [6–8]. Though promising results have been obtained, there continues to be a need to increase pDNA selectivity and yields while maintaining cost-effectiveness [9].

Traditional chromatographic methods based on porous beads had been reported for pDNA purification [3,10–13], but they suffer from low capacity, long separation times, small diffusion coefficients, and difficulties with scalability. The introduction of monoliths had overcome most of these drawbacks [14]. However other methods like precipitation [15,16] and extraction offer additional advantages such as technical simplicity and the possibility to use inexpensive equipment and chemicals. Extraction has been performed using both reverse micellar two-phase systems and aqueous two-phase systems (ATPSs). With the former systems it was possible to selectively extract RNA from a preconditioned bacterial lysate separating it from a model plasmid that remains in the aqueous phase [17]. Others experiments also show that this system could also separate plasmid isoforms.

Both polymer–polymer and polymer–salt ATPSs had been used for the downstream processing of plasmid DNA [18,19]. In either case, in addition to the aforementioned advantages can be added the mild conditions and non-toxicity of reagents offered by this methodology. Better results are usually obtained using a polymer–salt system with low-molecular weight poly(ethylene glycol) (PEG). Potassium phosphate, ammonium sulphate and sodium citrate had been used as salt with comparable results [19–21]. High recovery of pDNA is obtained in the salt phase with the elimination of most RNA, proteins and endotoxins to the polymer phase. However the system composition must be carefully chosen as significant differences between the top and bottom phase compositions, corresponding to high tie-line values, result in plasmid precipitation at the interface, with the concomitant yield reduction [21]. A recent study reported that while ammonium sulphate systems give higher purity than citrate systems, a lower recovery yield is obtained [22]. A compromise between these two parameters could be found using a mixture of the two salts.

In polymer–polymer systems the removal of RNA is not so efficient although 80% reduction of this contaminant has been

* Corresponding author. Tel.: +351 253604054; fax: +351 253678983.
E-mail address: jcmarcos@quimica.uminho.pt (J.C. Marcos).

0021-9673/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2008.07.095

achieved [18]. In addition, in these systems the lysate has to be previously desalted in order to precisely control the salt composition of the systems, which determines the partitioning of nucleic acids. In contrast, in salt–polymer systems the phase components could be directly added to the lysate [20].

Despite all the enumerated advantages of ATPSs, they have low selectivity when compared to chromatographic methods [19–21,23]. As a consequence they are used mainly to concentrate crude feedstocks and to remove major contaminants [18,21].

Increasing the specificity of ATPS purification processes is attractive because they can be used immediately after cell disruption to avoid the need for centrifugation or removal of cell debris. This could reduce downstream purification steps. Selectivity of ATPSs might be increased by employing ligands that bind specifically to the target biomolecules. Such approaches have been extensively explored in protein purification [24,25] but in the case of pDNA there is only one recent report of the utilisation of polyethylenimine to increase ATPS selectivity [6]. Potential pDNA ligands include metal ions [26], polymers [27], amino acids [28], proteins [29] and nucleic acids [15]. For proteins to act as pDNA ligands, they need to bind specific sequences of DNA within the plasmid. Zinc finger proteins are one of the best-known class of DNA-binding proteins and bind to their target DNA sequences with good sequence specificity via a well-understood molecular mechanism [30,31]. In a recent model it is proposed that these proteins bind initially non-specifically by electrostatic interaction and then run along the DNA strand until they find the recognition site where they strongly attached [32]. Indeed, the zinc finger protein used here has been shown previously to act as an effective affinity ligand for the isolation of pDNA [3,33]. The DNA sequence specificity of this zinc finger protein has been demonstrated previously [34,35] and is such that a plasmid that contains the target sequence may be isolated, while another plasmid that shares 7 base pairs (bp) of the 9 bp DNA recognition sequence is not [3]. In this experiment the zinc finger protein was fused to a glutathione-*S*-transferase (GST-ZnF) and immobilized on Sepharose.

The utilisation of GST-ZnF in ATPSs might therefore improve upon the selectivity observed when GST-ZnF is immobilised since the affinity interaction would take place in a homogeneous phase without the steric hindrance caused by the binding of GST-ZnF to a resin. In addition, the loss of product due to non-specific adsorption to the solid support could also be avoided.

The successful utilisation of GST-ZnF in ATPSs requires the selection of a two-phase system where it can partitioned into the phase with less contaminants. Polymer–polymer systems are preferred over polymer–salt systems since a high salt concentration may impair the interaction between the protein and pDNA. It has been reported that in some of these systems the majority of pDNA contaminants can accumulate in one particular phase. ATPSs have been described in which 97.5% of denaturated genomic DNA [36]; the majority of the total proteins [37], nearly 100% RNA and ~100% pDNA from bacterial cell lysate accumulate in the bottom dextran-rich phase [6].

It was hypothesised that partitioning of the DNA ligand to the PEG-rich phase could be accomplished by either (1) tailoring the two-phase PEG–dextran system such that the native GST-ZnF protein preferentially accumulates in the PEG phase, or (2) PEGylating the GST-ZnF protein to cause preferential partitioning into the PEG phase. The latter approach was investigated with conjugation of PEG to the fusion protein by reaction with either an amino or a thiol moiety on the protein. In this study the affinity isolation of pre-purified pDNA from model buffer solutions using native and PEGylated GST-ZnF was examined in PEG–dextran ATPSs.

2. Materials and methods

2.1. Reagents

PEGylation reagents mPEG-butyraldehyde (mPEG-ALD, molecular weight 20 kDa) and mPEG-maleimide (mPEG-MAL, molecular weight 5 kDa) were purchased from Nektar (San Carlos, CA, USA). Complementary synthetic oligonucleotides carrying the zinc finger recognition site 5′-TTT-TTT-TTT-TGG-GGC-GGC-TTT-TTT-TTT-TT-3′ and 5′-AAA-AAA-AAA-AAG-CCG-CCC-CAA-AAA-AAA-AA-3′ were purchased from Biomers (Ulm, Germany). They were annealed by combining 50 pmol of each oligonucleotide in a final concentration of 1 pmol/μL in water, heating to 95 °C and allowing to cool at room temperature. The resulting double strand synthetic probe was stored at −20 °C until further use. Lysozyme (LYZ), molecular weight 14.6 kDa; β-lactoglobulin (β-LAC), molecular weight 18.4 kDa; bovine serum albumin (BSA), molecular weight 66.0 kDa and GST, molecular weight 28.0 kDa were purchased from Sigma–Aldrich (St. Louis, MO, USA). PEG, molecular weight 600 and 1000 Da were obtained from Sigma. Dextran (DEX), molecular weigh 40 kDa (DEX 40) and 500 kDa (DEX 500) were both purchase from GE Healthcare (Chalfont St. Giles, UK).

2.2. Expression and purification of GST-ZnF protein

GST-ZnF fusion protein was expressed from the pGEX-2TK (GE Healthcare) vector in *Escherichia coli* BL21 cells and purified as described previously [3]. The resulting eluate containing GST-ZnF protein was dialysed at 7 °C against 50 mM Tris–HCl, 150 mM NaCl buffer pH 7.4. The purified protein was stored at −20 °C with 30% glycerol. When necessary, the protein was concentrated with a Vivaspin 6 concentrator column (Sartorius, Göttingen, Germany) and buffer exchanged using PD10 desalting columns (GE Healthcare). Protein concentration in solution was determined with the Bradford method (Pierce, Rockford, IL, USA) using BSA as standard.

2.3. Purification of pTS plasmid

The plasmid containing the ZnF binding sequence (pTS) was transformed into competent *E. coli* DH5α cells [38]. One single colony was used to inoculate 5 mL of LB (Luria-Bertani) medium supplemented with 50 μg/mL ampicilin. After 8 h culture, a 0.5-mL aliquot was use to inoculate 500 mL of similar media. The resulting culture was grown overnight at 37 °C and at 250 rpm. Cells were harvested by centrifugation and stored at −20 °C until needed. Plasmid DNA was purified from this cells after alkaline lysis, using a Maxi column (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

2.4. PEGylation and isolation of GST-ZnF conjugates

2.4.1. Amino selective PEGylation

Purified GST-ZnF protein (typically 0.5 mg/mL) was allowed to react with mPEG-ALD reagent at different molar ratios of mPEG to GST-ZnF (1:1, 5:1 and 25:1, i.e. 13, 65 and 325 nM mPEG, respectively) in 50 mM acetate buffer, pH 5.0 in the presence of 3 mg/mL of $NaCNBH_3$ at 7 °C (Table 1). The reaction was quenched with 0.5 M Tris–HCl, pH 7.4 (final concentration) after 24, 48 and 72 h. The protein-PEG products were analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) in a NuPAGE 4–12% Bis–Tris Gel (Invitrogen, Carlsbad, CA, USA) and stained with colloid blue stain. Removal of low-molecular weight reaction end-products and buffer exchange were performed using a PD10 desalting column (GE Healthcare) that had been pre-equilibrated with 50 mM Tris–HCl buffer pH 7.4. The PD10 eluate was then loaded into a

Table 1
Preparation and characterization of GST-ZnF protein PEGylation using mPEG-ALD (20 kDa) and mPEG-MAL (5 kDa) activated PEGs

Activated PEG	PEG equivalents	Reaction time (days)	PEGylation products (no.)	PEGylation efficiency (%)
mPEG-ALD	1	1	1	11.2
mPEG-ALD	5	1	2	47.5
mPEG-ALD	5	2	2	83.3
mPEG-ALD[a]	5	3	3	91.8
mPEG-ALD	25	3	8	95.3
mPEG-MAL	20	0.041 (1 h)	n.d	100

[a] Reaction chosen for the DNA affinity studies.

cation-exchange column (Hitrap FF, GE Healthcare) that has been pre-equilibrated with 50 mM Tris–HCl buffer pH 7.4 to remove any unconjugated PEG compounds. Size-exclusion chromatography (SEC) with a Hi-load Superdex 200TM column (GE Healthcare), eluted with 50 mM Tris–HCl buffer, pH 7.4 with 0.15 M sodium chloride at a flow rate of 1 mL/min was used to separate the protein-PEG conjugates from native protein. Protein components were detected at 280 nm. Reaction yield was estimated by comparative HPLC peak area analysis using Azur software (Kromatek, Great Dunmow, UK).

2.4.2. Sulfhydryl selective PEGylation

The GST-ZnF protein (typically 0.5 mg/mL, 0.5 mL) was first reduced with 100 mM DTT (dithiothreitol) (0.5 h, room temperature) and then the DTT was removed using a PD10 desalting column. The PEGylation reagent (13 μL), mPEG-maleimide (mPEG-MAL, 5 kDa; 100 mg/mL, 20 equivalents) in Tris–HCl buffer, pH 7.4 was then added to the reduced protein and the reaction mixture was incubated for 1 h at room temperature. PEGylated protein products were isolated from the reaction mixture using the same procedure described for the mPEG-ALD reaction. PEGylated peaks were collected and protein was concentrated using a Vivaspin 6 concentrator column.

2.5. Band shift assays

The apparent dissociation constants (K_d^{app}) of native and PEGylated GST-ZnF protein were determined by band shift assays using native polyacrylamide gel electrophoresis. In the reaction mixtures the amount of double-stranded oligonucleotide bearing the recognition sequence for the zinc finger protein was kept constant (7.43 pmol) while the protein concentration was varied from 0 to 2.28 μM (see Fig. 3 for details). DNA-binding reactions involved combining the zinc finger protein with the synthetic oligonucleotide in a solution of Tris–HCl buffer, pH 7.4, with 0.15 M sodium chloride, 20% glycerol in a total volume of 12 μL. The mixture was incubated for 30 min at room temperature and then resolved in a native Tris-glycine 10% gel (Invitrogen) using 1X Tris-glycine running buffer at 100 V for 80 min. Prior to the run, the gel was allowed to equilibrate in the same buffer for 2 h at 100 V. DNA was stained using ethidium bromide solution (0.5 μg/mL) and visualized by UV. After image acquisition a second staining was performed using colloid blue to detect protein.

To calculate the dissociation constants, the intensities of free DNA and DNA–protein complex bands in each lane were determined using Image J software [39]. The complex band intensity was taken as the sum of the intensities of all bands representing the association of DNA with one or more proteins (multimers).

The apparent dissociation constant of the native and PEGylated GST-ZnF protein was determined according to Wieland [40]. It was assumed that free protein (P_F) binds to free DNA (DNA_F) to form a complex P/DNA according to:

$$P_F + DNA_F \leftrightarrow \frac{P}{DNA} \quad (1)$$

Total concentration of DNA ($[DNA]_T$) and protein ($[P]_T$) is given by:

$$[DNA]_T = [DNA]_F + \left[\frac{P}{DNA}\right] \quad (2)$$

$$[P]_T = [P]_F + \left[\frac{P}{DNA}\right] \quad (3)$$

The dissociation constant:

$$K_d^{app} = [DNA]_F \frac{[P]_F}{[P/DNA]} \quad (4)$$

K_d^{app} can be expressed in terms of total concentration as:

$$K_d^{app} = \left([DNA]_T - \left[\frac{P}{DNA}\right]\right) \frac{([P]_T - [P/DNA])}{[P/DNA]} \quad (5)$$

For equal gel band intensities of free DNA and DNA complexed with protein we have:

$$[DNA]_F = \left[\frac{P}{DNA}\right] \quad (6)$$

Thus:

$$\left[\frac{P}{DNA}\right] = 0.5\,[DNA]_T \quad (7)$$

For this protein concentration $[P]_{50}$ we thus have:

$$K_d^{app} = [P]_{50} - 0.5[DNA]_T \quad (8)$$

For known $[DNA]_T$, $[P]_{50}$ was determined at 50% binding from the binding curve and used to calculate the apparent dissociation constant for the native GST-ZnF and PEGylated GST-ZnF proteins according to Eq. (8). The GST protein without the zinc finger fusion was used as negative control in similar binding conditions.

2.6. Phase diagrams

Two different PEG–DEX systems were chosen for the partition experiments. Binodal curves of PEG 600–DEX 40 and PEG 1000–DEX 500 were determined at 20 °C by titration according to Albertsson [37]. Briefly, small amounts of water were added to the biphasic systems of defined composition until a single homogenous phase was observed. At this point, the system composition was calculated and plotted to form the binodal curve.

2.7. Affinity-based isolation of pDNA by aqueous two-phase systems

Two systems composed of (A) 21.9% (w/w) PEG 600–11.7% (w/w) DEX 40 and (B) 15.7% (w/w) PEG 1000–8.9% (w/w) DEX 500 were prepared by weighing appropriate amounts of stock solutions of the phase forming polymers of different molecular weights in graduated tubes. Various concentrations of native or PEGylated GST-ZnF proteins were allowed to interact with a constant amount of pre-purified pTS plasmid (8.2 μg/4.5 pmol) in 20 mM Tris–HCl buffer for at least 10 min at room temperature in 0.2 mL (final volume). The resulting complex was then added to one of the two partitioning systems and pH was adjusted by adding 40 μL of 0.5M Tris–HCl, pH 7.4 buffer. Finally, ultra pure water was added to make 1 g total system weight. The partitioning systems were then mixed until all polymers were dissolved. Phase separation was accomplished by centrifugation (2000 × *g*, 2 min) and each phase was collected for further analysis. Phase turbidity was avoided by equilibrating at 30 °C prior to collection of each phase sample. All analyses were

conducted at room temperature on the same day that the partitioning systems were prepared.

2.8. Partition coefficient measurements

Quantitative determination of the protein content in each phase was performed by the Bradford assay using BSA as standard. The interference of PEG and DEX was overcome by diluting the samples 10-fold, and making a series of calibration curves using the diluted phase under analysis as solvent for the BSA standards. All concentrations were calculated using the average of at least three independent measurements. The partitioning of a protein in a two-phase system is described by the partition coefficient, K_c, and defined as $K_c = CT/CB$, where CT and CB are the concentrations of the protein in the top (PEG) and bottom (DEX) phases, respectively.

DNA quantification was performed in a similar way using the Picogreen assay (Invitrogen) in a 96-plate reader. Blanks of each system were prepared and a series of calibration curves using standard concentrations of pTS plasmid were made. Partition coefficients for pTS plasmid in each system were calculated using the same formula given above for proteins.

3. Results and discussion

3.1. PEGylation of GST-ZnF protein

The GST-ZnF protein has a molecular mass of 38.3 kDa with a chain length of 331 amino acid residues [33] of which 34 residues are lysine (Lys) and 10 are cysteine (Cys). Two different PEGylation conjugation strategies were thus envisaged (Fig. 1). Amino selective PEGylation was achieved using mPEG-ALD (20 kDa), which has a terminal aldehyde moiety capable of undergoing reductive amination with primary amines in the presence of sodium cyanoborohydride [41]. Often, a PEG-aldehyde reagent can be used to predominantly PEGylate the terminal amine in a protein, especially when the protein has an N-terminal methionine as is the case with GST-ZnF fusion protein used here [42].

Sulfydryl specific PEGylation was conducted using mPEG-MAL. Since thiol conjugation can be much more efficient than amine conjugation and there are 10 cysteines that are potentially available for thiol alkylation, a smaller molecular weight PEG reagent of 5 kDa was used. Four of the ten cysteines are located within the GST moiety and the other six within the ZnF protein. Three-dimensional structure analysis of the GST protein (1GTA, PDB Protein Data Bank) reveal that they are not cross-linked. The six cysteines within the zinc finger are coordinated to the zinc metal ion [34]. Therefore a number of different PEGylated products can result representing different sites and extent of conjugation. The relative amount of each PEGylated form in the product mixture would be expected to depend upon the rate of reaction at each site of conjugation. In respect to the PEGylation of therapeutic proteins, product heterogeneity results in varied biological activities. Although PEGylation heterogeneity with GST-ZnF may influence DNA binding, there would be likely a lower influence on partitioning properties, specially for highly PEG modified proteins [43].

PEGylation by reductive amination was conducted in acidic conditions (pH 5.0) with low (1) mPEG equivalents per protein in an attempt to form stable conjugates with only one PEG molecule per protein (mono-PEGylation). However, imine formation is inefficient and a relatively low PEGylation yield (11.2%, Table 1) was obtained when mono-PEGylation was desired. Analysis by SDS-PAGE of the reaction conducted with 5 equivalents mPEG-ALD (Fig. 2A, inset) indicated that multiple conjugations occurred (to yield native, mono-, di-, and tri-PEGylated GST-ZnF) as the reaction time was increased up to 72 h. Conjugate heterogeneity was also observed when the ratio of mPEG-ALD/GST-ZnF was increased to 25 equivalents (Table 1). The yield of different PEGylated forms was dependent on both the reaction time and the relative PEG molar ratio. As anticipated, attempts to increase PEGylation yield also resulted in increased molecular weight in the product heterogeneity. We found that PEGylation efficiency of up to 95.3% was achieved using 25 mPEG equivalents over a 72 h incubation period for reaction (resulting in at least eight different PEG-protein conjugates).

For the reactions that were conducted with 5 equivalents of mPEG-ALD per protein, at 24, 48 and 72 h, the PEG–GST-ZnF conjugates were isolated by first removing unconjugated PEG compounds by ion-exchange chromatography, and then fractionated native GST-ZnF by SEC as shown in Fig. 2A. The first peak displayed a retention time of 53 min and corresponds to high molecular weight multi-PEGylated conjugates. A second peak at 58 min, which was not baseline resolved with the peak at 53 min, was also observed. Electrophoresis analysis showed that this later peak contains mainly the mono-PEGylated conjugates (82.8%), although small amounts of high molecular weight species are also observed (data not shown). As the reaction time was increased from 24 to 72 h we observed a proportional increase of the 53 min peak area with a concomitant decrease of the 58 min peak area. The increase of the high molecular weight species was accompanied by an increase in PEGylation yield, as suggested by a decrease in the peak area corresponding to the native GST-ZnF protein (75 min).

To prepare the PEG–GST-ZnF conjugates for pDNA-affinity isolation studies in the aqueous two-phase system we selected PEGylation conditions of 72 h incubation with 5 equivalents mPEG-ALD/protein that gave an overall conversion of 92% of GST-ZnF. This selection involved a compromise between achieving relatively low product heterogeneity (the mono-, di- and tri-PEGylated population represented 67.1, 28.7 and 4.1%, respectively) and high reaction yield. The disadvantages of this compromise was that the different PEGylated forms may display different affinities with pDNA due to varying degrees of steric shielding of the pDNA binding site, as well as the possibility that the different forms may not partition equally to the PEG phase. Since separation of mono- from multi-PEGylated

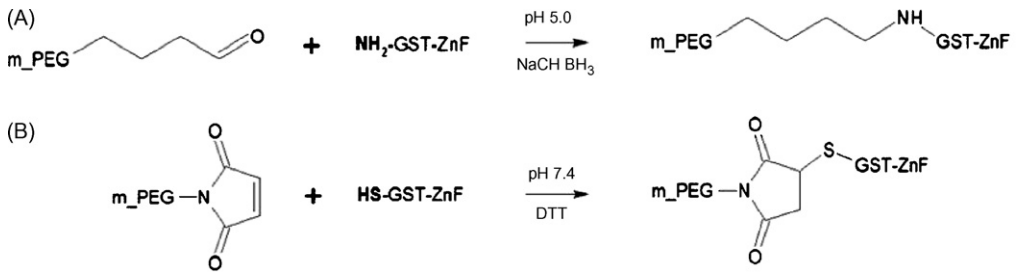


Fig. 1. Schematic representation of GST-ZnF PEGylation reactions. Highlighted with bold are the PEGylation targets for each PEGylation chemistry used. (A) mPEG-ALD (20 kDa) reaction and (B) mPEG-MAL (5 kDa) reaction. See Section 2 for complete reaction conditions.

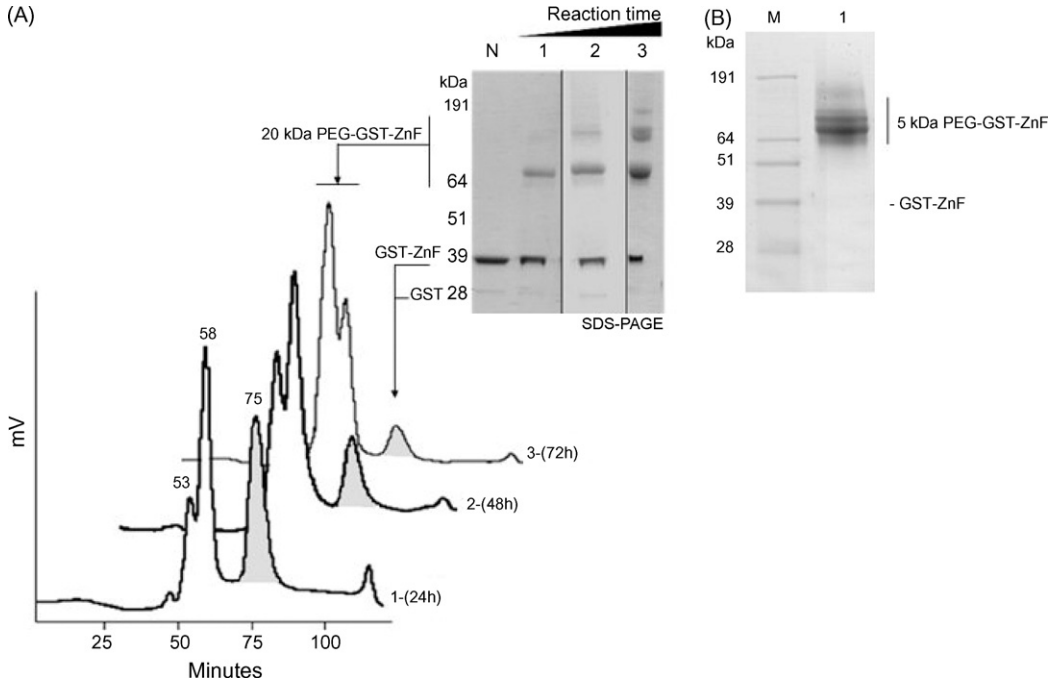


Fig. 2. Characterisation and isolation of PEG–GST-ZnF conjugates. (A) 5 equivalents mPEG-ALD per protein reactions. Size-exclusion chromatograms are shown after 24, 48 and 72 h reaction time. Peaks retention time for PEG conjugates are at 53 and 58 min and at 75 min for native GST-ZnF/GST proteins. SDS-PAGE inset shows the (N) Native GST-ZnF protein, (1) 24 h reaction, (2) 48 h reaction and (3) 72 h reaction. (B) SDS-PAGE characterisation of mPEG-MAL reaction. Lane (M) marker, (1) 1 h reaction.

species by SEC proved difficult the PEG–GST-ZnF conjugates from peaks at 53 and 58 min were collected and pooled as one sample for use in the DNA-affinity partitioning studies.

PEGylation of free cysteine residues in GST-ZnF protein was achieved using mPEG—5 kDa (mPEG-MAL). As shown in Fig. 1 mPEG-MAL undergoes an alkylation reaction with sulfhydryl groups to form a thioether bond. The GST-ZnF protein was incubated with 100 mM DTT prior to adding the mPEG-MAL to the protein solution. After removal of the DTT, the protein was incubated with PEG-MAL for 1 h and quantitative conjugation was observed (Table 1). Thiol alkylations are more efficient at mild pH values than amine alkylations. Interestingly with GST-ZnF, incomplete PEGylation reactions were observed with mPEG-MAL (40 equivalents) if the protein was not first incubated with DTT.

SDS-PAGE characterisation of the completed PEGylation reactions are shown in Fig. 2B. A narrow distribution of molecular weights was observed. This might be circumvented using higher molecular weight mPEG-MAL since we only used a 5-kDa reagent. Also decreasing the molar ratio of PEG-MAL per protein could be advantageous. Although we have seen that when using very low mPEG molar ratio (1 equivalent) a decrease of about 60% in the PEGylation efficiency was observed.

3.2. Apparent dissociation constant (K_d^{app}) determination

In order to evaluate the DNA binding affinity of the PEGylated GST-ZnF protein, gel-shift assays were performed to compare the apparent dissociation constants of native and PEGylated protein. Proteins at different concentrations were added to a constant amount of double-stranded oligonucleotide (7.43 pmol) bearing the recognition sequence for the zinc finger protein. Typical results for the mPEG-ALD modified GST-ZnF are shown in Fig. 3A and for the native GST-ZnF protein in Fig. 3B. With increasing protein concentration the intensities of bands representing free DNA decrease and those bands showing the bound complex increase. Quantitative analyses of the gels bands (from A and B, Fig. 3) are shown in Fig. 3C. Apparent dissociation constants were estimated from the protein concentration at which 50% DNA was bound using Eq. (8) (Section 2). Apparent dissociation constants of $K_d^{app} = 0.11\ \mu M$ for native GST-ZnF and $K_d^{app} = 0.24\ \mu M$ for aldehyde PEGylated GST-ZnF protein were obtained. In view of the known structural heterogeneity of the mPEG-ALD modified PEG conjugates the estimated value represents the average of the interaction of the three species with the pDNA molecule. Others have reported that in PEGylated proteins potential non-covalent interactions between PEG and other groups within the protein might lead to a region rich in this polymer [44,45]. If this region is near the active site then it may mask the pDNA binding site of the protein. The observed increase of K_d^{app} upon PEGylation corresponds to a change of 2 kJ/mol in the Gibbs free energy of binding, which is approximately the same energy as a single dipole–dipole interaction. This suggests that although a relatively high molecular weight mPEG-ALD (20 kDa) was used for the conjugation it does not significantly impair the binding of the protein with the oligonucleotide, possibly because PEG was attached to the protein in a region far from the binding site.

While the amino PEGylation approach resulted in only a slightly reduced DNA binding affinity as compared to that for the native protein, cysteine PEGylated GST-ZnF protein displayed no binding affinity with the double-stranded oligonucleotide. As shown in Fig. 3D (lanes 1–2), GST-ZnF that has been PEGylated using this approach did not bind to the oligonucleotide for any protein concentration tested (up to 3.8 μM). The ZnF protein is a consensus sequence zinc finger protein that binds to the sequence 5′-GGG-GCG-GCT-3′ belonging to the family Cys_2His_2 zinc finger motif [34]. Since in this case the PEG is bound to GST-ZnF by its cysteines residues it is possible that the PEGylation chemistry disrupted the coordination of the zinc ion to these amino acids, so modifying the binding site and impairing the interaction between the protein and the oligonucleotide. It is known that the structure of the zinc finger determines nucleotide sequence binding specificity [46].

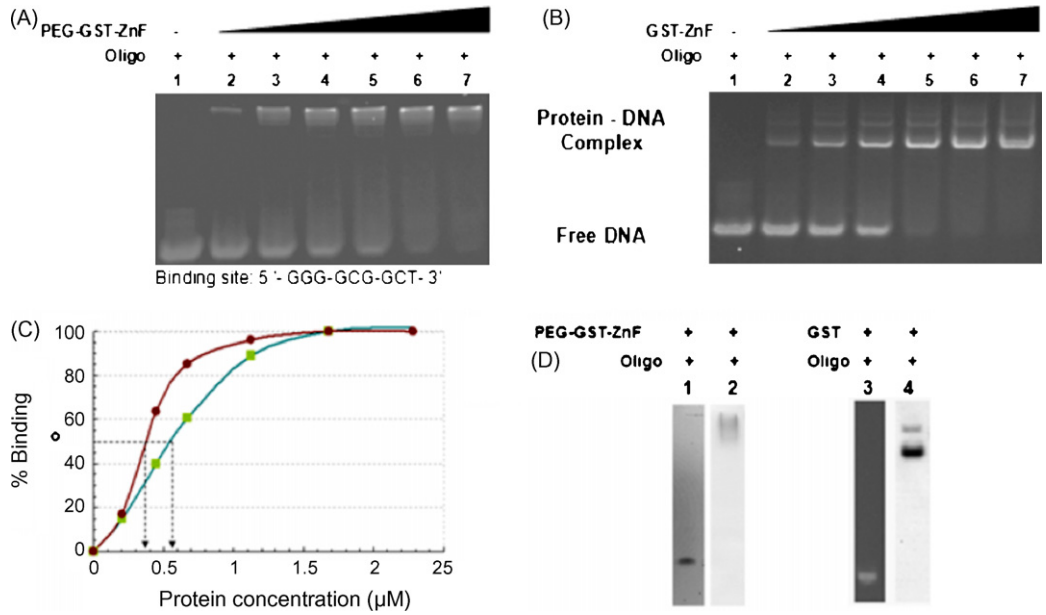


Fig. 3. Binding analysis of PEGylated and native GST-ZnF protein. (A) Binding of PEGylated (ALD) GST-ZnF (5 equivalents mPEG-ALD/protein reaction, 72 h reaction time). Lanes 1–7 increasing PEG–GST-ZnF concentrations: lane 1, 0.0 μM; lane 2, 0.20 μM; lane 3, 0.45 μM; lane 4, 0.67 μM; lane 5, 1.12 μM, lane 6, 1.68 μM; lane 7, 2.28 μM. Below the gel the DNA binding sequence for the zinc finger protein is shown. (B) Binding of native GST-ZnF. Lanes 1–7 increasing PEG–GST-ZnF concentrations: same concentrations as in the gel A. (C) Chart representation of the data of the gels shown in (A) and (B). Circles, binding values for GST-ZnF. Squares, binding values for PEG–GST-ZnF (ALD). Dashed arrows, $[Protein]_{50}$, protein concentration for 50% binding. (D) Lanes 1–2 binding of PEG–GST-ZnF, 3.8 μM (20 equivalents mPEG-MAL/protein reaction). (1) Ethidium bromide stain; (2) colloid blue stain. Lanes 3–4 binding of GST, 5.7 μM (negative control). (3) Ethidium bromide stain; (4) colloid blue stain.

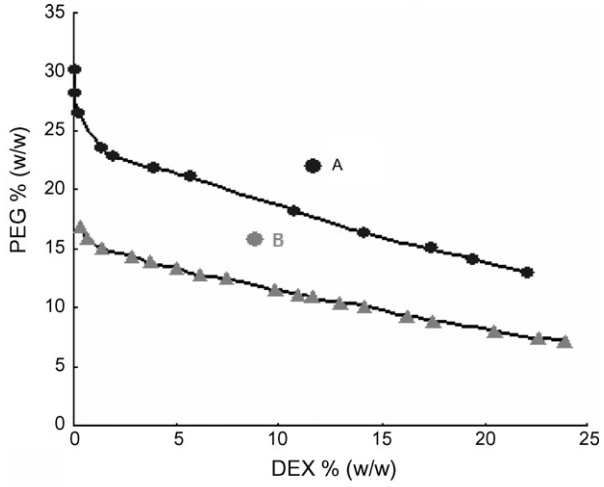


Fig. 4. Phase diagrams with binodals for PEG 600–DEX 40 (circles) and PEG 1000–DEX 500 (triangles), at room temperature. Two systems were selected from the two-phase region of each binodal: (A) 21.9% (w/w) PEG 600–11.7% (w/w) DEX 40 and (B) 15.7% (w/w) PEG 1000–8.9% (w/w) DEX 500.

3.3. Protein partitioning in aqueous two-phase systems

Successful affinity partitioning requires that the affinity ligand directs the partition of the target biomolecule to the phase where less contaminants accumulate. Adequate systems will be the ones where most contaminants and the target biomolecule accumulates in one phase and the affinity ligand in the other. In this way the affinity ligand will promote the extraction of the target biomolecule to the phase with less contaminants increasing its purity. With this goal in mind two different ATPSs were chosen (Fig. 4) that differ mainly with respect to the molecular weight of the phase forming polymers used. In system A, PEG 600 and DEX 40 were used while in system B, PEG 1000–DEX 500 were used. The phase composition (% w/w of each phase forming polymers) is also different in both systems.

The protein partition coefficients (K_c) in both systems A and B are summarised in Table 2. In system A GST-ZnF protein accumulated in the bottom phase ($K_c = 0.013$) while in system B it accumulated in the PEG-rich top phase ($K_c = 37.89$). Additionally, in system A PEG–GST-ZnF partitioned mainly to the PEG phase with an enhanced partition coefficient of $K_c = 17.19$. This corresponds to 97.5% of the PEG–GST-ZnF protein accumulating in the top phase of system A, while 96.5% of native GST-ZnF accumulated in the bottom phase. Harris et al. [47] claimed similar partitioning behaviour for an antibody against alkaline phosphatase that had been chemically modified (38%) with mPEG-ALD, with 99% partitioning of the antibody conjugate to the top PEG phase while unmodified antibody partitioned mainly (90%) to the bottom phase (in a system consisting of 6% PEG 8000–8% DEX 40, 150 mM NaCl, 10 mM sodium phosphate, pH 7.2).

Table 2
Comparison of different proteins partition coefficient (K_c) on the systems (A) 21.9% (w/w) PEG 600–11.7% (w/w) DEX 40 and (B) 15.7% (w/w) PEG 1000–8.9% (w/w) DEX500 in Tris–HCl buffer, pH 7.4

System	PEG–GST-ZnF[a]	BSA (66.0 kDa)	GST-ZnF (39.1 kDa)	β-LAC (18.4 kDa)	LYZ (14.6 kDa)
A PEG 600–DEX 40	17.19	0.012	0.013	0.21	2.55
B PEG 1000–DEX 500	n.d.	46.16	37.89	6.91	4.81

Total volume, 1 mL.
[a] 72 h aldehyde PEGylation reaction (5 equivalents mPEG/protein).

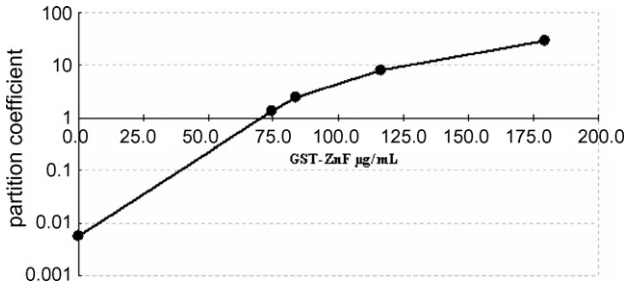


Fig. 5. Partition coefficients of pTS plasmid DNA versus GST-ZnF concentration in the top phase of the system B, 15.7% (w/w) PEG 1000–8.9% (w/w) DEX 500 at pH 7.4.

Table 3
Analysis of pTS partition with increase PEG–GST-ZnF concentration (μg/mL) in the top phase of the system A, 21.9% (w/w) PEG 600–11.7% (w/w) DEX 40 at pH 7.4

Systems	Top phase		pTS partition coefficient (K_c)
	[PEG–GST-ZnF] ± S.D.	% pTS ± S.D.	
A1	0	0.10 ± 0.04	0.0003
A2	38.5 ± 2.8	62.1 ± 4.6	0.883
A3	52.8 ± 7.8	72.9 ± 3.5	1.282
A4	244.2 ± 4.3	82.1 ± 2.5	2.435
A5	293.4 ± 5.6	100.0 ± 1.3	TP[a]

[a] TP—top phase.

There is an approximately linear relationship between the logarithm of the partition coefficient in ATPS and the degree (n) of PEGylation (number of mPEG molecules per protein). However, this relationship, predicted on the assumption that each grafted PEG contributed equally to K_c, did not hold over the larger range of n [43]. The PEG-protein partition behaviour is governed by the phase systems used, surface density, topographical and conformation distribution of their grafted PEGs and PEGylation chemistry used [43,48].

In order to assess the feasibility of using these systems in purification protocols it was studied the partition behaviour of three model proteins; BSA (MW 66.0 kDa, pI (isoelectric point) = 4.8); β-LAC (MW 18.4 kDa, pI = 5.1) and LYZ (MW = 14.6 kDa, pI = 11). From Table 2 it can be seen that as the molecular weight of the protein was increased its partition coefficient also increased in system A but decreased in system B. It seems that in these systems conditions the protein molecular weight is an important determinant of its partitioning behaviour rather then the protein charge. From system A to B the dextran molecular weight was changed from 40,000 Da to 500,000 Da (12.5 times higher) whereas similar PEG molecular weights (600 and 1000 Da, respectively) were used in both systems. This behaviour is usually attributed to an excluded volume effect and has been observed for several other proteins [49,50]. Increasing the molecular weight of one polymer steers the protein to the other phase, an effect that increases with the molecular weight of the protein.

With systems in which high molecular weight dextran is used it is likely that similar proteins from *E. coli* will accumulate in the same phase system as the ligand. It is known that the average molecular weight of proteins from *E. coli* is around 40 kDa [51], a very close molecular weight to the affinity protein used here (38.3 kDa). Though, other factors such as hydrophobicity, salt composition, polymers molecular weight, temperature and pH might influence the partitioning of biomolecules in PEG–dextran systems [37,52]. Thus, it is essential to study the behaviour of the total protein partitioning from a complex mixture of a bacterial cell lysate in these polymer–polymer systems to precisely select an appropriated ATPS for the use of affinity ligands. The utilization of PEGylated GST-ZnF conjugates rather than native protein in systems with low-molecular weight dextran would confer an advantage since the majority of the proteins from a bacterial lysate would likely accumulate in the bottom phase and the PEG–GST-ZnF in the top.

3.4. Affinity partition of pDNA by native and PEGylated GST-ZnF protein

The partitioning of pTS plasmid as a function of added native GST-ZnF ligand in system B, pH 7.4 is shown in Fig. 5. In this system native protein accumulated mainly in the top phase, K_c = 37.89, while in the absence of ligand pTS plasmid DNA accumulated in the bottom phase, K_c = 0.005. A notable shift of pTS partition (~5900-fold) was observed upon increasing the GST-ZnF concentration in the top phase, demonstrating that the protein interacts with pDNA in ATPSs and changes its partitioning to the top phase (up to K_c (pDNA) = 29.42, Fig. 5). As the concentration of the ligand was increased a shift of the true partition coefficient of pTS towards that of the ligand, K_c = 37.89 was observed.

We similarly examined the system PEG 600–DEX 40 (A) to evaluate the pTS affinity partition using the PEG–GST-ZnF conjugates (Table 3). In these systems pTS accumulated almost totally in the bottom phase, K_c = 0.0003 while the PEGylated ligand accumulated 97.5% in the top phase (96.5% native GST-ZnF protein accumulates on the bottom). For PEG-ligand concentrations in the top phase up to 293 μg/mL pTS accumulated completely (100%) in this phase (the detection limit of the Picogreen assay is 0.2 ng dsDNA (double-stranded DNA) and 4.5 pmol (8.2 μg) of pTS DNA was added to each system, from which it is estimated that less than 0.002% pTS DNA might be in the bottom phase). Fig. 6 shows a comparison of pTS and GST-ZnF protein partitioning with and without PEGylation.

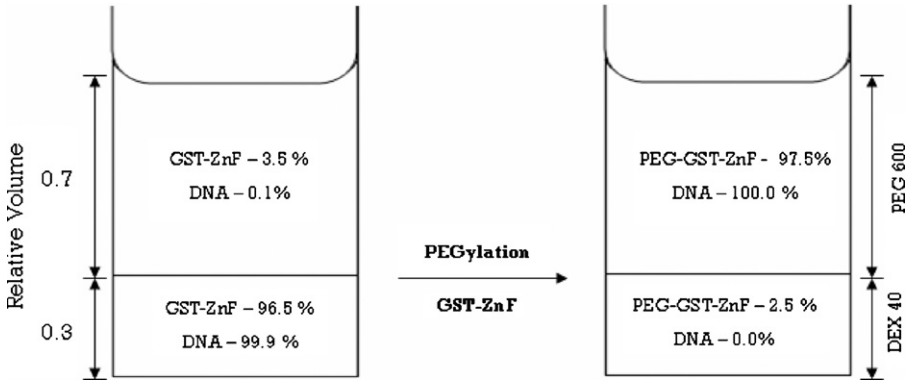


Fig. 6. Schematic representation of the pTS DNA and GST-ZnF partition before and after the protein ligand be PEGylated with mPEG-ALD. System A, 21.9% (w/w) PEG 600–11.7% (w/w) DEX 40 at pH 7.4.

Using either native or PEGylated GST-ZnF ligand a shift on the pTS DNA partition was observed in each system. However, in system A when the PEG-ligand concentration in the top phase was 244.2 μg/mL it had a lower pTS partition coefficient (K_c = 2.43) compared to the pTS partition coefficient of K_c of 29.42 obtained when 179.42 μg/mL of native protein was used in system B (Fig. 5). This may be due to different composition of the systems that influence the partition of the GST-ZnF/pTS complex. It could also be that although the dissociation constant is similar for the PEGylated and native protein in buffer solution, as determined by the gel-shift assays (Fig. 3), it may be different in the PEG phase. One reason for this is that the protein bound PEG could interact strongly with the free PEG in solution making a PEG-rich region around the protein that impairs the interaction with the DNA molecule. In either case it seems that adding more ligand overcame these problems as total accumulation of pTS in the top phase was achieved using 293.4 μg/mL of ligand.

4. Conclusions

It has been shown that both PEGylated and native GST-ZnF protein were able to significantly enhance pDNA partition to the PEG-rich phase, though in ATPSs of different composition. The utilisation of PEGylated ligands rather than native protein in affinity purification of pDNA confers advantages since the ligand might be easily recovered following elution by a second extraction of the pTS partitioned from the PEG phase to a new salt phase. The PEGylated ligand would accumulate in the PEG phase and both could be reused. Additionally, ligands bearing affinity tags to specific supports such as GST or His tags will facilitate its recovery. Polymer–polymer ATPSs would be preferential as the primary system as this would circumvent the utilization of a salt as the main phase forming component, thus avoiding possible interferences of the salts in the protein ligand–DNA interactions.

ATPSs have been recognized as a good process alternative for the separation and purification of pDNA. Affinity partitioning can enhance the previously low selectivity of these systems as the pDNA has a high and selective affinity for the protein ligand. Additionally, the pDNA/ligand complex accumulates in a low contaminant content phase (top phase) rather than the bottom. A simple and fast method for the isolation pDNA from the dextran-rich phase to the PEG-rich phase with no pDNA loss in the bottom phase has been demonstrated here by exploiting PEGylated GST-ZnF. This approach for pDNA purification opens good perspectives to the design of a procedure for a high selectivity isolation of pDNA.

Acknowledgment

H.B. acknowledges a Ph.D. grant (SFRH/BD/16296/2004/XS49) from FCT, the Portuguese Foundation for Science and Technology.

References

[1] P.A. Shamlou, Biotechnol. Appl. Biochem. 37 (2003) 207.
[2] D.M. Prazeres, G.N. Ferreira, G.A. Monteiro, C.L. Cooney, J.M. Cabral, Trends Biotechnol. 17 (1999) 169.
[3] J. Woodgate, D. Palfrey, D.A. Nagel, A.V. Hine, N.K. Slater, Biotechnol. Bioeng. 79 (2002) 450.
[4] L.M. Sandberg, A. Bjurling, P. Busson, J. Vasi, R. Lemmens, J. Biotechnol. 109 (2004) 193.
[5] R.A. Darby, G.M. Forde, N.K. Slater, A.V. Hine, Biotechnol. Bioeng. 98 (2007) 1103.
[6] S.P. Duarte, A.G. Fortes, D.M. Prazeres, J.C. Marcos, J. Chromatogr. A 1164 (2007) 105.
[7] R.J. Lander, M.A. Winters, F.J. Meacle, B.C. Buckland, A.L. Lee, Biotechnol. Bioeng. 79 (2002) 776.
[8] J.C. Murphy, J.A. Wibbenmeyer, G.E. Fox, R.C. Willson, Nat. Biotechnol. 17 (1999) 822.
[9] M.M. Diogo, J.A. Queiroz, D.M. Prazeres, J. Chromatogr. A 1069 (2005) 3.
[10] G.N. Ferreira, J.M. Cabral, D.M. Prazeres, Biotechnol. Prog. 16 (2000) 416.
[11] M.M. Diogo, J.A. Queiroz, G.A. Monteiro, S.A. Martins, G.N. Ferreira, D.M. Prazeres, Biotechnol. Bioeng. 68 (2000) 576.
[12] R. Lemmens, U. Olsson, T. Nyhammar, J. Stadler, J. Chromatogr. B 784 (2003) 291.
[13] L.Z. Li, Y. Liu, M.S. Sun, Y.M. Shao, J. Chromatogr. A 1139 (2007) 228.
[14] A. Zochling, R. Hahn, K. Ahrer, J. Urthaler, A. Jungbauer, J. Sep. Sci. 27 (2004) 819.
[15] J. Kostal, A. Mulchandani, W. Chen, Biotechnol. Bioeng. 85 (2004) 293.
[16] M.D. Costioli, I. Fisch, F. Garret-Flaudy, F. Hilbrig, R. Freitag, Biotechnol. Bioeng. 81 (2003) 535.
[17] N. Streitner, C. Voss, E. Flaschel, J. Biotechnol. 131 (2007) 188.
[18] C. Kepka, J. Rhodin, R. Lemmens, F. Tjerneld, P.E. Gustavsson, J. Chromatogr. A 1024 (2004) 95.
[19] S.C. Ribeiro, G.A. Monteiro, J.M. Cabral, D.M. Prazeres, Biotechnol. Bioeng. 78 (2002) 376.
[20] A. Frerix, M. Muller, M.R. Kula, J. Hubbuch, Biotechnol. Appl. Biochem. 42 (2005) 57.
[21] I.P. Trindade, M.M. Diogo, D.M. Prazeres, J.C. Marcos, J. Chromatogr. A 1082 (2005) 176.
[22] G.A. Gomes, A.M. Azevedo, M.R. Aires-Barros, D.M. Prazeres, Sep. Purif. Technol., doi:10.1016/j.seppur.2008.01.026, in press.
[23] A. Frerix, P. Geilenkirchen, M. Muller, M.R. Kula, J. Hubbuch, Biotechnol. Bioeng. 96 (2007) 57.
[24] G. Johansson, Methods Mol. Biol. 147 (2000) 105.
[25] B.Y. Zaslavsky, Aqueous Two-Phase Partitioning: Physical Chemistry and Bioanalytical Applications, Marcel Dekker, New York, 1995.
[26] L. Tan, W.B. Lai, C.T. Lee, D.S. Kim, W.S. Choe, J. Chromatogr. A 1141 (2007) 226.
[27] D. Costiolia, R. Freitaga, F. Hilbrig, React. Funct. Polym. 65 (2005) 351.
[28] F. Sousa, S. Freitas, A.R. Azzoni, D.M. Prazeres, J. Queiroz, Biotechnol. Appl. Biochem. 45 (2006) 131.
[29] R.A. Darby, A.V. Hine, FASEB J. 19 (2005) 801.
[30] A. Klug, J. Mol. Biol. 293 (1999) 215.
[31] N.P. Pavletich, C.O. Pabo, Science 252 (1991) 809.
[32] J.H. Laity, B.M. Lee, P.E. Wright, Curr. Opin. Struct. Biol. 11 (2001) 39.
[33] S. Ghose, G.M. Forde, N.K. Slater, Biotechnol. Prog. 20 (2004) 841.
[34] J.R. Desjarlais, J.M. Berg, Proc. Natl. Acad. Sci. U.S.A. 90 (1993) 2256.
[35] Z.R. Zhang, M.D. Hughes, L.J. Morgan, A.F. Santos, A.V. Hine, Biotechniques 35 (2003) 988.
[36] R. Ohlsson, C.C. Hentschel, J.G. Williams, Nucl. Acids Res. 5 (1978) 583.
[37] P.Å. Albertsson, Partition of Cell Particles and Macromolecules, Wiley, New York, 1986.
[38] J. Sambrook, E. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[39] W.S. Rasband, ImageJ, US National Institutes of Health, Bethesda, MD, 1997.
[40] G. Wieland, P. Hemmerich, M. Koch, T. Stoyan, J. Hegemann, S. Diekmann, Nucl. Acids Res. 29 (2001) 1054.
[41] M.J. Roberts, M.D. Bentley, J.M. Harris, Adv. Drug Deliv. Rev. 54 (2002) 459.
[42] O. Kinstler, G. Molineux, M. Treuheit, D. Ladd, C. Gegg, Adv. Drug Deliv. Rev. 54 (2002) 477.
[43] C. Delgado, M. Malmsten, J.M. Van Alstine, J. Chromatogr. B 692 (1997) 263.
[44] E.L. Furness, A. Ross, T.P. Davis, G.C. King, Biomaterials 19 (1998) 1361.
[45] B.A. Kerwin, B.S. Chang, C.V. Gegg, M. Gonnelli, T. Li, G.B. Strambini, Protein Sci. 11 (2002) 1825.
[46] C. Branden, J. Tooze, Introduction to Protein Structure, Garland Pub., New York, 1991.
[47] J.M. Harris, M.R. Sedaghat-Herati (University of Alabama), Preparation and use of polyethylene glycol propionaldehyde—US Pat. 5,252,714 (1993).
[48] C. Delgado, F. Malik, B. Selisko, D. Fisher, G.E. Francis, J. Biochem. Biophys. Methods 29 (1994) 237.
[49] S. Sasakawa, H. Walter, Biochemistry 11 (1972) 2760.
[50] N. Dalloraa, J. Klemza, P. Filho, Biochem. Eng. J. 34 (2007) 92.
[51] S. Sundararaj, A. Guo, B. Habibi-Nazhad, M. Rouani, P. Stothard, M. Ellison, D.S. Wishart, Nucl. Acids Res. 32 (2004) D293.
[52] B.A. Andrews, A.S. Schmidt, J.A. Asenjo, Biotechnol. Bioeng. 90 (2005) 380.

Chapter 4

Dual Affinity Method for Plasmid DNA Purification in Aqueous Two-Phase Systems

The results presented in this chapter were submitted for publication as follow:

- H.S.C. Barbosa, A. V. Hine, S. Brocchini, N.K.S. Slater and J.C. Marcos (2008), Dual Affinitty Method for Plasmid DNA Purification in Aqueous Two-Phase Systems. (Submitted to Journal of Chromatography A).

The results described in this chapter were presented in the following conference:

- H.S.C. Barbosa *et al.*, Dual Affinity Method for Plasmid DNA purification in Aqueous Two-Phase systems, Affinity, Iceland, 2009 (Poster presentation)

4.1 Abstract

The DNA binding fusion protein, LacI-His_6-GFP, together with the conjugate PEG-IDA-Cu(II) (10 kDa) was evaluated as a dual affinity system for the pUC19 plasmid extraction from an alkaline bacterial cell lysate in poly(ethylene glycol) (PEG) /dextran (DEX) Aqueous Two-Phase Systems (ATPS). In a PEG 600 – DEX 40 ATPS containing 0.273 nmol of LacI fusion protein and 0.14 % (w/w) of the functionalised PEG-IDA-Cu(II), more than 72 % of the plasmid DNA partitioned to the PEG phase, without RNA or genomic DNA contamination as evaluated by agarose gel electrophoresis. In a second extraction stage, the elution of pDNA from the LacI binding complex proved difficult using either dextran or phosphate buffer as second phase, though more than 75 % of the overall protein was removed in both systems. A maximum recovery of approximately 27 % of the pCU19 plasmid was achieved using the PEG-dextran system as a second extraction system, with 80-90 % of pDNA partitioning to the bottom phase. This represents about 7.4 µg of pDNA extracted per 1 mL of pUC19 desalted lysate.

Keywords: Aqueous Two-Phase Systems; Plasmid Purification; Affinity ligand; Immobilized Metal-ion Affinity Partitioning; Lac repressor; bacterial cell lysate

4.2 Introduction

Non-viral vectors, such as plasmid DNA, have been shown to be adequate to mediate both gene therapy and DNA vaccination. Expression of therapeutic proteins has been attained both *in vitro* and *in vivo* using these vectors [1]. Also, they are able to stimulate humoral and cellular immune responses to a specific antigen [1; 2]. The increased number of clinical trials using naked/plasmid DNA (there are currently 281 ongoing trials) have brought a focus on the large-scale manufacturing of pDNA as high demand for these vectors is soon expected [3].

The most common methods for large-scale purification of plasmids are chromatographic [4]. However, other methods such as liquid-phase extraction in Aqueous Two-Phase Systems (ATPSs) may be interesting alternatives since the problem of access and mass transfer within solid-phase chromatography matrices for large plasmids is avoided. Several studies show that ATPS extraction has several other advantages, such as; operational simplicity, easy scale-up and the possibility to combine several process steps in a single operation [5; 6; 7]. However, one of the main drawbacks of this technique is its low selectivity, which may be overcome by the use of affinity ligands [8; 9]. The specificity and bio-recognition properties of affinity ligands for pDNA capture from a crude feedstock with large and structurally related impurities (such as, high molecular weight RNA, endotoxin and shear genomic DNA) may facilitate the preparation of pDNA products that fulfil the stringent specifications for market approval.

Although affinity based separations have been extensively applied for protein extraction in ATPS's [10; 11; 12], only two recent reports refer to the utilisation of pDNA ligands [8; 9]. Duarte *et al*. [8] have used the cationic polymer poly(ethyleimine) (PEI) as non-specific ligand for the purification of pDNA from crude cell lysates in a two-step ATPS process. Although the authors report up to 100 % pDNA recovery in the form of DNA/PEI polyplexes, with only minor protein impurities, they did not elute the pDNA from the polyplexes. Although it

is not yet clear if the pDNA could be eluted and used in pure form, concerns remain about the use of PEI based vectors for gene therapy applications since the cationic polymers might present some toxicity issues [13]. In addition, the PEI/DNA interactions are mainly electrostatic, thus there is a potential risk for non-specific binding of the cationic polymer to other negative charged molecules, such as endotoxins, RNA and genomic DNA, which would also raise potential safety concerns.

Alternatively, Barbosa *et al.* [9] have described a protein-based ligand composed of Glutathione-S-Transferase protein (GST) fused to a Zinc Finger transcription factor (ZnF) designed as GST-ZnF. The GST-ZnF affinity protein was able to isolate pDNA bearing a ZnF recognition site from the bottom to the top phase of a PEG-dextran ATPS in either the native or a PEGylated form. GST-ZnF binds to a specific base sequence at the DNA recognition site, from which it is not easily eluted. Thus, no attempts were made to extract pure pDNA from a complex mixture of a crude cell lysate.

In both approaches the ligand has to be covalently bound to poly(ethylene glycol) (PEG); through PEGylation [14; 15]. In the PEG-dextran systems used, this PEGylation drives the ligand to the PEG phase where less contaminants usually accumulate. The presence of the ligand in the PEG phase may draw the pDNA to this phase by affinity binding, so separating it from the contaminants.

Though effective, this approach is limited by the need for PEGylation. This can be a complex and inefficient procedure and several products are often obtained with different degrees of PEGylation [9]. Such inhomogeneity in the PEGylated ligands may require further purification of the final DNA sample to obtain a homogeneous product and this in turn may lead to high production costs and to the increase in batch operation times. Additionally, because many PEGylations are inefficient, a stoichiometric excess of the reagent is needed, which can potentially make the entire process uneconomic for some ATPS purifications.

To circumvent these problems a modified PEG might be used that binds site specifically and *in situ* with the ligand. This approach has been already explored

for the affinity purification of proteins bearing a His-tag in aqueous two-phase systems [16; 17]. Thus, Immobilized Metal-ion Affinity Partitioning (IMAP) has been used as an alternative to Immobilized Metal-ion Affinity Chromatography (IMAC)[18]. IMAP exploits the grafting of a chelator, such as iminodiacetic acid (IDA), onto one the phase forming polymers, usually the PEG polymer. The IDA ligand chelates a transition metal ion (e.g. Cu^{2+} or Ni^{2+}), which binds preferentially to electron-rich amino acid residues such as histidines and cysteines that are accessible on the surface of proteins [19]. This enables the extraction of specific his-tagged proteins into the PEG phase while the bulk of contaminant proteins without a his-tag remain in the other phase [17].

The strategy set out above does not require the previous PEGylation of the protein and the binding to the His-tag occurs almost instantly. Here, the use of this highly selective method to enhance the partitioning of a DNA affinity protein to the PEG-rich phase of a PEG-dextran ATPS is exploited in order to steer the target pDNA to the same phase.

One of the ligands that is being investigated for the affinity chromatographic purification of plasmid DNA is the LacI repressor protein fused to both a His_6 tag and a Green Fluorescent Protein (GFP) moiety [20; 21]. As shown in Figure 1, the LacI-His_6-GFP fusion protein is a tetramer protein (two dimer units) composed of three different domains: 1) A LacI domain that binds sequence specifically the *Lac*O_3/*Lac*O_1 operator on the pUC19 plasmids; 2) The GFP domain that enables the specific detection of the protein in a mixture of other proteins; 3) The His_6 tag that bridges the two domains and enables the purification of the protein by IMAC. This protein has been immobilised onto an IMAC column for the purification of pDNA directly from crude bacterial cell lysates, yielding highly pure supercoiled pDNA with no RNA, genomic DNA or denatured pDNA [20; 21].

Here, we report a dual affinity ATPS approach for the purification of plasmid DNA in which the heterofunctional DNA binding ligand LacI-His_6-GFP is combined with an Immobilized Metal–Ion Affinity PEG. The functionalised PEG

binds to the His-tag of this protein, resulting in the Immobilized Metal-Ion Affinity Partitioning (IMAP) of the LacI-His_6-GFP / pDNA complex to the PEG-rich top phase. The advantages of using this approach are that the covalent PEGylation of the DNA ligand is not necessary. The high pDNA sequence specificity of the DNA binding protein is effectively exploited and the binding of the functionalised PEG to the DNA binding protein can be performed in a crude cell extract that is then mixed with the pDNA lysate, thus potentially reducing operating costs and processing complexity.

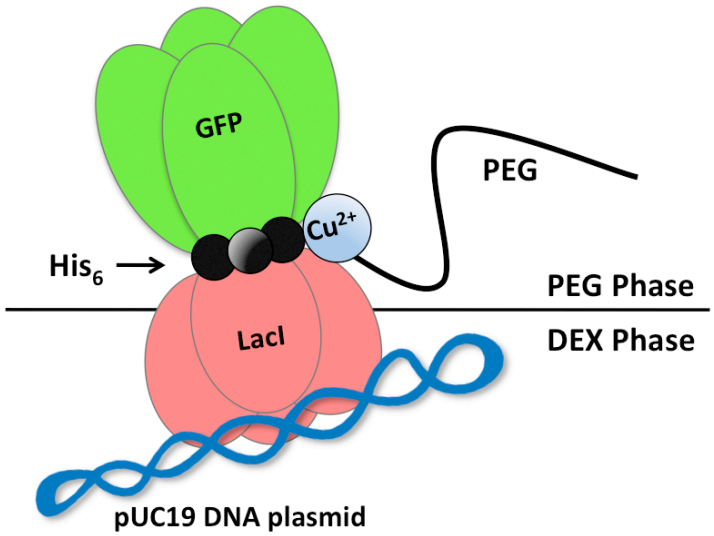


Figure 1 - Schematic representation of the dual affinity method developed for pDNA purification using PEG/dextran aqueous two-phase systems.

4.3 Material and Methods

4.3.1 Reagents

PEG polymers, MW 10,000 and 600 Da were obtained from Sigma-Aldrich (St. Louis, MO, USA) and dextran MW 40 kDa was purchased from GE Healthcare (Chalfont St Giles, UK). LB broth media, β-lactoglobulin protein, boron trifluoride ethyletherate (BF_3) and epichlorohydrin were purchased from Sigma-Aldrich.

4.3.2 Preparation of PEG-IDA-Cu(II)

PEG-IDA-Cu(II) was synthesised using the epichlorohydrin method, as similarly described in [22]. Briefly, 30 g of PEG 10,000 Da was dissolved in 100 mL toluene and 0.3 mL BF_3 (50 % w/w) and 0.4 mL epichlorohydrin were added dropwise. The reaction mixture was stored for 48h and 2ml NaOH (45 %) was then added to the mixture, which was slowly stirred for 4h at 37 ºC. PEG was precipitated and the powder was then dissolved in 100 mL carbonate buffer pH 12.1 containing 0.35 M iminodiacetic acid (IDA) and the mixture was stirred at 70 ºC overnight. After three extractions with chloroform, the product was crystallised twice in ice-cold diethyl ether (yielding 85-90 % of total PEG). The Cu(II) complex with PEG-IDA was obtained by dissolving the PEG-IDA product in 50 mL sodium acetate buffer pH 4 containing 5g of $CuSO_4$. The extraction was repeated yielding proximally 75 % of the original PEG. The 0.88 µmoles of Cu(II) per µmol PEG was determined by atomic absorption spectrometry using appropriate standards.

4.3.3 Expression and purification of LacI-His_6-GFP protein

LacI-His_6-GFP DNA construct [20] was transformed into *E. coli* BL21-gold cells (Stratagene, La Jolla, CA) and a single fresh colony was inoculated into 200 mL of LB media containing 50 µg/mL ampicillin and incubated overnight at 30 ºC and 200 rpm. 1:1000 dilution of this culture was then prepared in 0.5 L of similar media and the LacI-His_6-GFP was expressed constitutively overnight at 200 rpm and 30 ºC. Cells were then harvested by centrifugation and the pellet

was stored at -80 ºC until needed. Protein was purified using a 1 mL Histrap HP column (GE Healthcare). Briefly, a cell pellet from 0.5 L cell culture was re-suspended in 20 mL of 20 mM phosphate, 500 mM Nacl and 40 mM Imidazole buffer pH 7.4 and lysed by 3×3 second pulse sonication at maximum power inside an ice bucket. The resulting lysate was clarified by centrifugation at 19,000 rpm for 15 min at 4 ºC and loaded into a Histrap column using a syringe pump. The column was extensively washed after the loading step with the re-suspended buffer until no absorbance was detected by UV 280 nm. Protein was eluted using 20 mM phosphate, 500 mM NaCl, 500 mM imidazole buffer pH 7.4. The resulting eluate containing LacI-His_6-GFP was dialysed at 7 ºC against PBS buffer pH 7.4 to remove imidazole. The buffer was exchanged when necessary by using a PD10 desalting column (GE Healthcare). Protein concentration in solution was determined with the Bradford method (Pierce, Rockford, IL, USA) using BSA as standard.

4.3.4 Preparation of pure pUC19 plasmid and pUC19 alkaline lysate.

Escherichia coli DH5-α strain, harbouring the pUC19 plasmid was cultivated and the pUC19 plasmid was purified as previously described [9]. Preparation of the alkaline cell lysate was performed according to Duarte *et al.* [8]. The resulted clarified lysate was filtered using a 0.22 μm syringe filter and stored at -20 ºC until needed. On the day of the experiment the buffer was exchanged to 20 mM phosphate buffer pH 7.4 using a PD10 column.

4.3.5 Affinity partitioning in aqueous two-phase systems

Aqueous two-phase systems composed of PEG 600 – DEX 40 were prepared by weighing appropriate amounts of stock solutions (30 - 40 % w/w) of the phase forming polymers in 2 mL graduated tube. Immobilized metal-ion affinity partitioning experiments were performed by replacing part of the PEG with the modified PEG-IDA-Cu(II) (the amount of PEG derivative is given hereafter as a percentage of the total mass of PEG present in the systems). The two proteins under study (LacI-His_6-GFP and β-Lactoglobulin) were then added into the systems and mixed by vortex shaking. The crude cell lysate containing pUC19 or the pure pUC19 solution was added and pH was adjusted to the desired pH

using concentrate stock buffers. Finally, ultra pure water was added to make 1.5 g total system weight. The systems were well mixed by vortex shaking until all polymers were completely dissolved. Phase separation was accomplished by placing the tubes in a water bath at 30 ºC for ~2 hours to avoid phase turbidity. Each phase was collected and all further analyses were conducted at room temperature on the same day that the partitioning systems were prepared.

The second extraction stage was performed by collecting the desired top phase of a system (0.5mL) and adding a fresh lower phase (DEX 40, 14 % w/w) or 0.4 mL potassium phosphate solution, K_2HPO_4-KH_2PO_4 (40 %). IPTG or NaCl was added to the final concentrations described in Table 4.

4.3.6 Analytical Methods

4.3.6.1 LacI-His_6-GFP and total protein Partition analysis

The quantitative determination of the total protein in each phase was performed by the Bradford assay using BSA as standard. The specific concentration of LacI-His_6-GFP fusion protein in each phase was estimated by measuring the GFP fluorescence in a 96 well plate reader calibrated against pre-purified LacI-His_6-GFP as described previously [23]. All concentrations were calculated using the average of at least 3 independent measurements.

4.3.6.2 DNA partitioning analysis

Total DNA quantification was performed using a Picogreen assay (Invitrogen, Carlsbad, CA, USA) in a 96 well plate reader. Blanks for each system were prepared and a series of calibration curves using standard concentrations of purified pUC19 plasmid in the blank top or bottom phase of each system was used. To diminish any quenching effects of LacI-His_6-GFP bound to pUC19 plasmid all samples were treated with 1µg/µl DNase-free proteinase K (Sigma) for 1.5h prior to readings.

The partitioning of a component, x, in a two-phase system is described by the partition coefficient, K_x, and calculated as $K_x = C_{TP}/C_{BP}$, where C_{TP} and C_{BP} are the concentrations of the protein in the top and bottom phases, respectively.

The partition percentage (PP) of each component, in each phase (i) from the total recovered component was calculated as:

$$PP = \frac{[x]_i \times V_i}{x_{top} + x_{bottom}}$$

and the final pDNA yield as

$$yield = \frac{[x]_i \times V_i}{x_{added}}$$

4.4 Results and Discussion

4.4.1 LacI-His$_6$-GFP protein affinity partitioning

Whether the modified PEG-IDA-Cu(II) would increase the free LacI-His$_6$-GFP partitioning to the top phase of the 18 % (w/w) PEG 600 – 14 % (w/w) DEX 40 system was first examined. Several ATPSs were made with increasing concentration of functionalised PEG, but keeping the LacI-His$_6$-GFP protein concentration constant (Fig. 2A).

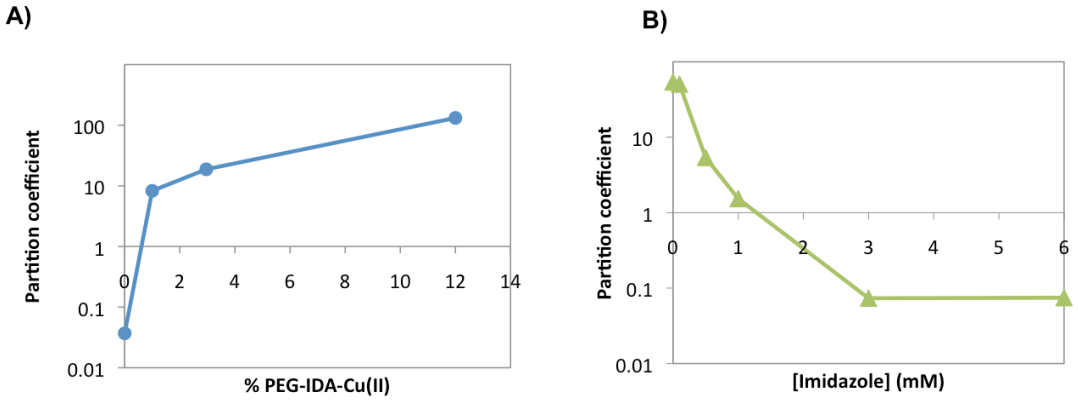


Figure 2 - Effect of modified PEG-IDA-Cu(II) and imidazole on the LacI-His$_6$-GFP partition coefficient (K_{LacI}) in systems composed by 18 % (w/w) PEG 600 – 14 % (w/w) DEX 40 in 50 mM Tris-HCl pH 7.4. A) Variation of K_{LacI} with the addition of PEG-IDA-Cu(II). B) Variation of K_{LacI} with the addition of imidazole in the presence of 10 % PEG-IDA-Cu(II).

In the absence of PEG-IDA-Cu(II) the LacI-His$_6$-GFP protein partitioned to the dextran-rich bottom phase (K_{lacI}= 0.037, 92.5 %). The partitioning behaviour of LacI-His$_6$-GFP was significantly enhanced by increasing the concentration of PEG-IDA-Cu(II) and it accumulated to 99.5 % in the PEG-rich top phase (K_{lacI}= 131) when 12 % of total PEG was replaced by the functionalised PEG.

To confirm that the binding between protein/PEG was due to the coordination between the transition metal ion (Cu(II)) chelated by PEG-IDA and the imidazole rings of the histidines on the LacI-His$_6$-GFP surface, we added increased concentrations of exogeneous imidazole in solution. As seen in Fig. 2B, in systems where 10 % of the PEG was replaced by PEG-IDA-Cu(II), increased

concentration of exogeneous imidazole steered the LacI-His_6-GFP protein to the bottom phase. For 3 mM imidazole the protein partitioning coefficient was K_{lacI}= 0.073, which is close to that observed when no functionalised PEG was added (K_{lacI}= 0.037). This confirms that IMAP of the LacI fusion protein was achieved.

4.4.2 pUC19 plasmid affinity partition

Fig. 3 shows the effect of increased LacI fusion protein concentration on systems with constant amount of pUC19 (7.9 pmol) and PEG-IDA-Cu(II) (12 % (w/w)). In systems with no LacI fusion protein present, the pre-purified pUC19 plasmid partitioned to the bottom phase (96.9 %) whereas the his-tagged LacI fusion protein partitioned almost totally to the top phase in systems without the pDNA (see Fig. 2). When the LacI-His_6-GFP protein (0.062 nmol) was combined with pUC19 DNA in the same system there was a slight increase in pUC19 partitioning to the top phase (K_{pUC19} increased from 0.027 to 0.129). When the LacI-His_6-GFP concentration was increased further, a decrease in pUC19 partition coefficient was observed, eventually to such an extent that no DNA was detected in the top phase of these systems.

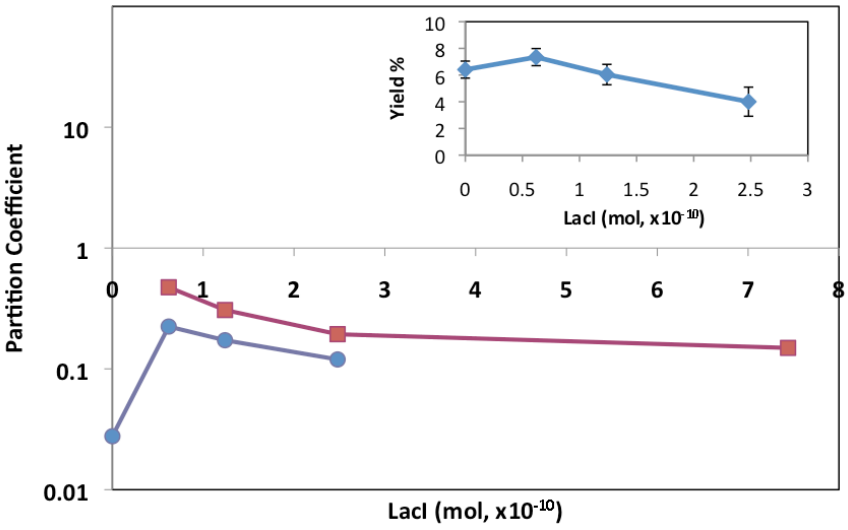


Figure 3 - pUC19 (circles) and LacI-His_6-GFP (squares) partition coefficient (*K*) in systems composed with increased LacI concentration and constant amount of PEG-IDA-Cu(II). Systems composed by 18.0 % (w/w) PEG 600 – 13.4 % (w/w) DEX 40 with 12 % of the PEG replaced by PEG-IDA-Cu(II) in 50 mM Tris-HCl pH 7.4. Each system contained constant amount of pUC19 (7.9 x10^{-12} mol, 14.36 µg). pUC19 accumulates on bottom phase only for systems with 7.4x10^{-10} mol of LacI-His_6-GFP. Inset shows the DNA yield on the top phase.

The top phase recovery of pUC19 in all these systems was around 4-6 % as shown in the inset to Fig. 3. Agarose gel electrophoresis analysis of the bottom and top phases of the systems was performed to determine if the pDNA was present in its free form or bound to the LacI fusion protein (Fig. 4). The pDNA was found to be bound to the affinity protein as an increase in the pDNA molecular weight was observed in the bottom phase of these systems, particularly for systems with higher protein concentration as 0.24 or 0.74 nmol (lane 8 and 9 respectively). This initially suggested that the affinity between the DNA-protein was higher than that between protein and PEG-IDA-Cu(II), so compromising the use of the PEG-IDA-Cu(II) to enhance the partitioning of pDNA/protein complex to the top phase.

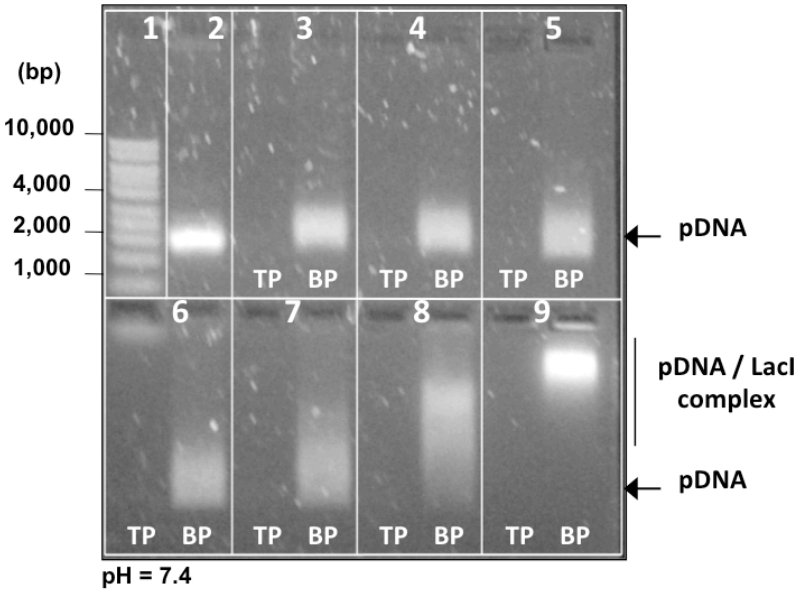


Figure 4 - DNA electrophoresis analysis of the top and bottom phase of systems composed with increased LacI-His$_6$-GFP concentration with constant amount of pUC19 plasmid (7.9 x10^{-12} mol, 14.36 µg) and PEG-IDA-Cu(II) (12 % of the total PEG). Systems composed by 18 % (w/w) PEG 600 – 13.4 % (w/w) DEX 40, 50 mM Tris-HCl pH 7.4. 1) DNA marker; 2) pUC19 plasmid; 3) Top phase (TP) and bottom phase (BP) analysis of system with only pUC19 plasmid; 4) System with pUC19 plasmid and 12 % PEG-IDA-Cu(II); 5-9 systems with pUC19 plasmid and 12 % PEG-IDA-Cu(II) and 0.3, 0.6, 1.2, 2.4, 7.4 x10^{-10} mol of LacI-His$_6$-GFP protein, respectively.

To further determine why the pDNA/LacI complex had partitioned to the bottom phase of these systems we also measured the decrease of the PEG-IDA-Cu(II) ligand in these systems, as shown in Fig. 5 and Table 1. Employing lower PEG-IDA-Cu(II) concentrations, while keeping pUC19 (2ρmol, 3.6 µg) and LacI (190

ρmol, 54.2μg) concentrations constant, enhanced the partitioning of both to the top phase (Fig. 5A). The percentage of LacI-His_6-GFP and pUC19 partitioned to the top phase in systems without PEG-IDA-Cu(II) were 27.2 % and 11.4 % respectively (Table 1). By contrast, with similar conditions but with 0.15 % (w/w) of the total PEG replaced with the functionalised PEG, the LacI fusion protein partitioned to the top phase.

Under these conditions, the LacI protein accumulated 85 % in the top phase (K_{lacI}= 5.09) while pUC19 partitioned 89.4 % (K_{pUC19}= 7.29) into the same phase. However, as the concentration of PEG-IDA-Cu(II) was increased, the partition coefficients of both protein and pDNA fell. Agarose gel analysis of the top and bottom phases of these systems (Fig. 5B) now showed that the pDNA was complexed with LacI protein in the top phase as a shift in the molecular weight of the supercoiled and open-circular bands of pDNA was observed (see for reference Fig. 5B, lane 1).

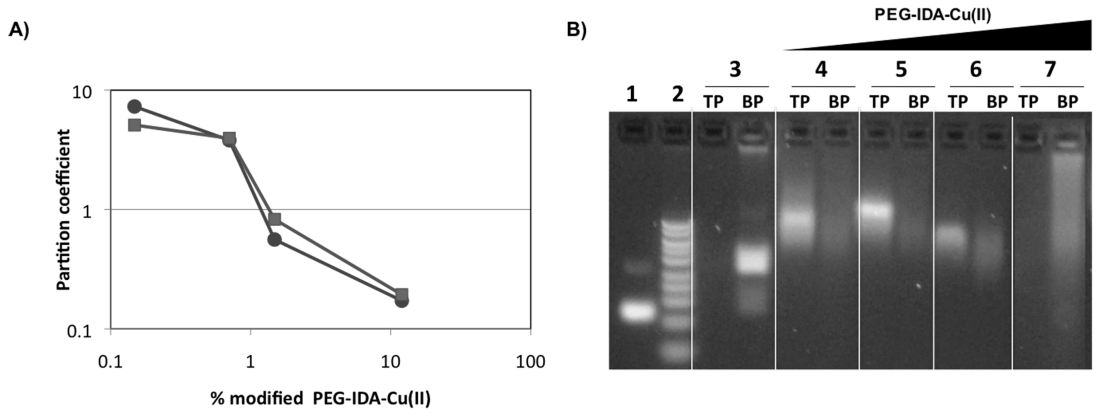


Figure 5 – Effect of PEG-IDA-Cu(II) on the pUC19 plasmid and LacI-His_6-GFP protein partition in systems composed by 16.0 % (w/w) PEG 600 – 14.3 % (w/w) DEX 40 in 15 mM phosphate buffer pH 7.4. A) Partition coefficient of pUC19 (circles) and LacI-His_6-GFP (squares) in each system contained constant amount of pUC19 (2 x 10^{-12} mol, 3.6 μg) and LacI-His_6-GFP (1.9 x 10^{-10}, 54.2 μg). B) DNA gel electrophoresis of top and bottom phase of systems described in A; 1) pUC19 plasmid (control); 2) DNA marker (see Fig. 4); 3) Top phase (TP) and bottom phase (BP) of system composed with pUC19 and PEG-IDA-Cu(II) (12 % total PEG); 4-7) 0.15, 0.71, 1.5, 12 % of modified PEG added, respectively.

It is speculated that the retention of the pDNA in the bottom phase for higher PEG-IDA-Cu(II) concentrations may have arisen because the functionalised PEG significantly changed the physical properties of the top phase, possibly due to

the formation of dimers or multimeric PEG-IDA-Cu(II) chains. Since a 10 kDa PEG-IDA-Cu(II) was used, it is possible that a very high MW chain was formed, particularly at high concentrations of the functionalised polymer. Thus, excluding from that PEG-rich phase the pDNA/protein complex. This behaviour, an excluded volume effect, has been observed for several other proteins in ATPS with increase of the molecular weight of one of the phase forming polymers [24; 25].

Table 1 - Top phase partition percentage of pUC19 plasmid and LacI-His$_6$-GFP protein with increased PEG-IDA-Cu(II) concentration in systems composed by 16.0 % (w/w) PEG 600 – 14.3 % (w/w) DEX 40, 15 mM phosphate buffer pH 7.4. Each system contained constant amount of pUC19 (2 x 10^{-12} mol, 3.6µg) and LacI-His$_6$-GFP (1.9 x 10^{-10} mol, 54.2 µg).

PEG-IDA-Cu(II) % (w/w)	Partition percentage, % mass	
	LacI-His$_6$-GFP	pUC19
0	27.2 ± 2.5	11.4 ± 3.5
0.15	85.4 ± 0.7	89.4 ± 5.3
0.71	75.9 ± 0.5	75.3 ± 6.1
1.48	59.5 ± 0.6	49.9 ± 1.6
12.00	31.9 ± 0.9	21.6 ± 2.7

The selective partitioning of the his-tagged LacI protein was confirmed by replacing it by β-lactoglobulin (Table 2). In the same polymer- polymer systems in phosphate buffer pH 7.4 the β-lactoglobulin protein partitioned to the top phase, $K_{\beta\text{-lacto}}$= 1.5, and a slight increase of its partition value $K_{\beta\text{-lacto}}$= 1.8 was observed when 0.148 % of the modified PEG-IDA-Cu(II) was added. When both β-lactoglobolin and pUC19 were added in the same systems there was no significant shift of pDNA partitioning to the top phase (K_{pUC19}= 0.015) similar to systems where no protein was added (K_{pUC19}= 0.024).

4.4.3 Extraction of pUC19 plasmid from crude cell lysate

Affinity capture of pUC19 directly from crude cell lysate was performed by adding the desalted lysate (14 % load, w/w) into systems with 0.148 % of the

modified PEG. Preliminary experiments with the original pUC19 alkaline lysates had shown that no pDNA/LacI interaction was observed (data not shown). However, it was observed that exchanging the buffer prior to extraction the binding between the pDNA/LacI occurred. Therefore, in all further experiments, the lysate was desalted (i.e., buffer exchanged to phosphate buffer pH 7.4) prior to extraction. This observation may have been due to the possible excess of sodium dodecyl sulfate (SDS) detergent used in the cell lysis and to the low pH ≈ 5.6 of the lysate. SDS is a known protein denaturant and the optimal pH for LacI/pDNA binding occurs at pH 7.4 [20; 21].

Table 2 - Partition coefficient of pUC19 plasmid in the presence of β-lactoglobulin protein (negative control) in systems composed by 18.0 % (w/w) PEG 600 – 13.4 % (w/w) DEX 40 in 15 mM phosphate buffer pH 7.4.

ATPS	β–Lactoglobulin mol x 10^{-9}/μg	pUC19 mol x 10^{-12}/μg	PEG-IDA-Cu^{2+} % (w/w)	Partition Coefficient		DNA Partition percentage (%)[a)]
				β-lactoglobulin	pUC19	
1	-	3.62 / 6.3	-	-	0.024	3.954 ± 0.95
2	-	3.62 / 6.3	0.148	-	0.007	0.984 ± 0.05
3	3.5 / 65.3	3.62 / 6.3	0.148	2.0	0.015	2.46 ± 0.05
4	3.5 / 65.3	-	0.148	1.8	-	-
5	3.5 / 65.3	-	-	1.5	-	-

a) DNA partition percentage to the top phase of each system

Although no genomic DNA (gDNA) is thought to be present on the lysates (see lane 3, Fig. 6A or B), we will report to total DNA partition (K_{DNA}) instead of pUC19 partitioning when lysates are used since the quantification method used here (Picogreen) does not differentiate from both forms of DNA.

As shown in Table 3, without LacI-His_6-GFP or PEG-IDA-Cu(II) ligands total protein ($K_{protein}$= 0.78) and DNA (K_{DNA}=0.022) from the desalted lysate partition to the bottom phase of these systems in phosphate buffer pH= 7.4. When 0.148 % (w/w) of immobilized metal-PEG was added to the system an increase of total

protein partitioning from the pUC19 lysate (to $K_{protein}$ = 0.96) was observed, showing that some non-specific binding might occur. The competitive binding of the modified PEG to other proteins is expected to be low in ATPSs that have low concentrations of a binding competitor added, such as imidazole.

Increase of the his-tagged LacI concentration in these systems resulted in a concomitant enhancement of the DNA partitioning to the top phase. For example, in systems with 0.27 nmol of LacI-His_6-GFP, the DNA partition coefficient increased to K_{DNA}=1.54 representing 72.8 % top phase partition percentage whereas, in systems without LacI protein and PEG-IDA-Cu(II) the DNA partitioned only 3.6 % to this phase (K_{DNA}= 0.022).

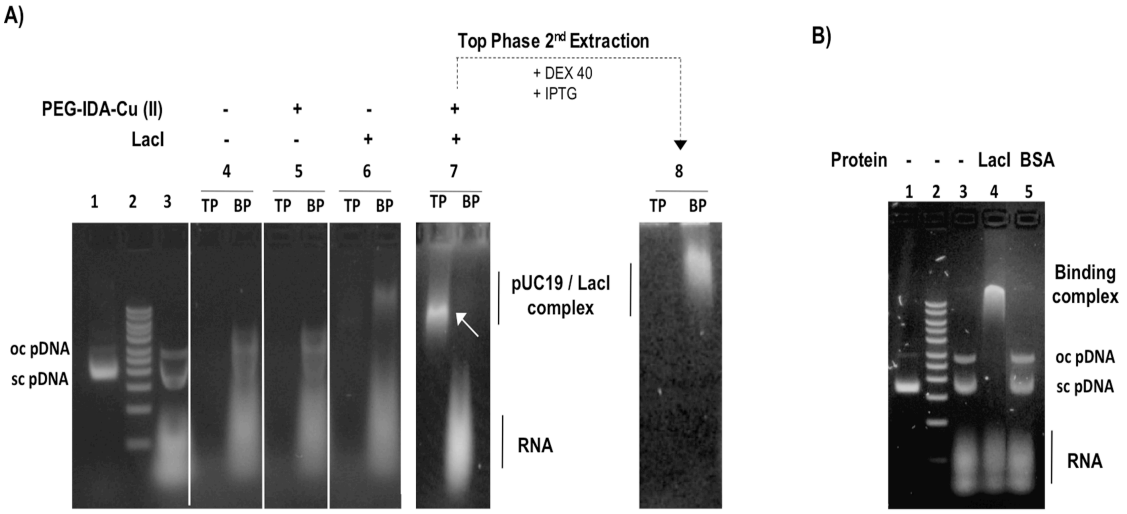


Figure 6 - Agarose gel analysis of the interaction between LacI-His_6-GFP protein and bacterial lysate in ATPS (A) and in solution (B). A) Top phase (TP) and bottom phase (BP) of systems composed with bacterial lysate and LacI fusion protein. Systems composed by 16 % (w/w) PEG 600 – 13.3 % (w/w) DEX 40 (0.148 % of the PEG was replaced with PEG-IDA-Cu(II)); 1) pUC19 plasmid; 2) Marker; 3) bacterial lysate containing pUC19; 4) top and bottom phase system with only bacterial lysate; 5) system with bacterial lysate and PEG-IDA-Cu(II); 6) Systems with pUC19 and LacI-His_6-GFP protein (without modified PEG); 7) system with bacterial lysate, $2.73x10^{-10}$ mol LacI fusion protein and 0.14 % of PEG-IDA-Cu(II); 8) back extraction of system shown in lane 7 to a new DEX 40, 2.5 mM IPTG. B) Analysis of pUC19 lysate with LacI protein and BSA in solution. 1) pUC19; 2) marker; 3) bacterial lysate; 4) bacterial lysate and LacI-His_6-GFP; 5) bacterial lysate and BSA.

The top phase recovery (yield) achieved with the above system composition was about 49.2 % of total DNA added. Given the crude nature of cell lysates it is significant that neither RNA nor genomic DNA is co-extracted to the top phase of these systems (lane 7-TB, Fig. 6A). Fig. 6B shows the binding of the LacI

protein to the pUC19 plasmid in solution. As shown for ATPS, the LacI fusion protein binds preferentially to the pDNA instead of RNA and no interaction is observed with BSA. Although no genomic DNA is present in this pUC19 lysates, as no high MW bands in the gels are present, it is also known that the LacI fusion protein binds almost 4 x 10^6 fold more strongly to its operator compared to other non-specific sites within the DNA sequence [20]. Although the his-tagged LacI might bind to one operator it requires two operators, *Lac*O3/*Lac*O1in the pUC19 plasmid, in order for the protein/DNA complex for optimal binding.

Here it was accomplished one step elimination of RNA and genomic DNA, two major pDNA contaminants. For the therapeutically use of plasmid DNA a lack of contamination by RNA, genomic DNA, host proteins, and endotoxin is necessary to meat the stringent specifications of regulatory agencies [26].

4.4.4 Second extraction / pUC19 elution

To elute the pUC19 plasmid from its LacI-His_6-GFP affinity protein and remove excess of the co-purified host proteins, a second extraction stage was performed with either DEX 40 polymer or a potassium phosphate salt buffer as the second phase. The top phase of ATPS with composition shown on system 6 (Table 3) was collected and, additionally to the phase forming polymer/salt, IPTG or NaCl were also added. IPTG is known to induce an allosteric change in the LacI protein, allowing the release of the plasmid, whereas NaCl addition increases the ionic strength of the solution and weakens the electrostatic interactions that mediate the binding of the protein to the DNA [20].

As seen in Table 4, the use of dextran as the second phase resulted in the accumulation of the pDNA in the bottom phase, K_{DNA}= 0.09, for systems with 2.5 mM IPTG / 0.3M NaCl, and K_{DNA}=0.04 when only IPTG was added. Total protein partitioned to the opposite PEG phase with $K_{protein}$= 2.91 and $K_{protein}$ = 2.63, respectively for each system, whereas in both systems LacI partitioned to the bottom with the DNA.

Table 3 - Effect of LacI-His$_6$-GFP increase concentration on the pUC19 plasmid partition from a desalted bacterial cell lysate. System composed by 18 % (w/w) PEG 600 – 13.3 % (w/w) DEX 40 in 15mM phosphate buffer pH 7.8.

ATPS	PEG-IDA-Cu(II) % (w/w)	LacI x10^{-10} mol	Partition coefficient (*K*)			DNA PP top[a)] (%)	DNA Yield (%)
			Total protein	LacI	pUC19		
1	-	-	0.78	-	0.022	3.6 ± 0.2	1.7 ± 0.1
2	0.14	-	0.96	-	BP[b)]	BP[b)]	BP[b)]
3	-	1.36	0.99	0.77	0.14	17.4 ± 1.2	8.0 ± 0.4
4	0.14	0.82	1.32	0.72	0.13	16.5 ± 1.8	12.2 ± 0.6
5	0.14	1.36	3.45	1.24	1.00	59.8 ± 1.6	36.7 ± 1.2
6	0.14	2.73	4.17	2.35	1.54	72.8 ± 4.3	49.2 ± 3.1

a) PP Top – Partition Percentage on Top Phase
b) BP – Bottom Phase only
pUC19 lysate: 109.6 µg/mL protein and 11.9 µg/mL DNA

Table 4 - Second extractions of pUC19 plasmid using the top phase of the system 16 % (w/w) PEG 600 – 13.3 % (w/w) DEX 40 composed with bacterial cell lysate, 2.73 x10^{-10} mol LacI-His$_6$-GFP protein and 0.14 % of PEG-IDA-Cu(II) in,15mM phosphate buffer pH 7.8.

Second phase	IPTG (mM)	NaCl (M)	Partition coefficient			Protein removal (%)	DNA PP [a)] (%)	DNA yield (%)
			Total protein	LacI	DNA			
DEX40 (pH 7.4)	2.5	0.3	2.91	0.71	0.09	76.3 ± 0.79 (bottom)	81.8 ± 3.9 (bottom)	27.3 ± 1.3 (bottom)
	2.5	-	2.63	0.51	0.04	85.9 ± 1.6 (bottom)	89.2 ± 8.4 (bottom)	25.5 ± 2.7 (bottom)
K_2HPO_4 – KH_2PO_4 (pH 7.4)	1.2	-	0.31	TB[b)]	12.86	74.4 ± 2.6 (top)	94.7 ± 5.9 (top)	23.8 ± 1.5 (top)
	-	-	0.33	TB[b)]	2.90	76.7 ± 0.8 (top)	81.3 ± 2.2 (top)	9.0 ± 0.3 (top)

a) DNA PP – DNA Partition Percentage on Top Phase
b) TB- Top phase only

Agarose gel analysis of the bottom phase of the second extraction system showed that the pDNA remained in the bound form (Fig. 6A, lane 8-BP) and thus elution had not been properly achieved. Further attempts to elute the DNA were made by adding PEG-IDA-Cu(II) to the systems or by increasing the IPTG concentration but no improvement to the pDNA elution was observed. In both cases more than 75 % of total protein was removed (85 % for back-extraction systems without NaCl). Much greater protein removal would have been achieved if the pUC19 plasmid had been eluted from the LacI protein since under these conditions the LacI protein would partition to the PEG-IDA-Cu(II) rich phase, as shown in Fig. 2A.

We speculated that polymer–salt systems might improve upon the pDNA elution since a high salt concentration may impair the pDNA / protein interaction. As shown in Table 4, however, similar results were obtained and no elution was achieved with K_2HPO_4 - KH_2PO_4 solution. With salt as the second phase, the DNA and LacI partitioned to the PEG-rich phase rather than to the salt phase where total protein accumulated ($K_{protein} \approx 0.3$). A slightly improvement on the final pDNA yield from 9.0 % to 23.8 % was attained when 1.2 mM IPTG was added but in both cases the pDNA was still in the bound form.

The lack of DNA elution by IPTG is confusing and requires further study. It is conceivable that steric hindrance by PEG may impede the access of IPTG to the binding site on LacI and thus not induce the conformational change in the 3D structure of LacI. A similar effect would result if IPTG partitioned to the opposite phase to the LacI/pDNA complex or if the protein/pDNA complex aggregated during the extraction process, hindering access to the site for IPTG binding. These hypotheses are being investigated and the protocol is being optimised for pUC19 elution.

Nonetheless with this current protocol, 75 % of total protein was removed without either genomic DNA nor RNA contamination. Although an overall yield of ~25 % was obtained, this represents about 1.4 mg of pDNA per 200 mL

desalted lysate which is about a 9-fold improvement compared to that reported for LacI-His_6-GFP immobilized into an IMAC column [21].

4.5 Conclusions

A dual affinity partitioning strategy is presented for the purification of pUC19 in PEG-dextran ATPS. A system composed of PEG-IDA-Cu(II)/ LacI-His_6-GFP has been shown to be effective for the affinity purification of pDNA from bacterial alkaline lysates without the need for covalent PEGylation of the ligand. This represents an advantage over this process since there is no need to perform the expensive and time consuming PEGylation reactions or the subsequent PEG-ligand purification steps. Although pDNA elution from the complex proves to be difficult we believe that with further improvements this method may be appropriate for the development of large-scale manufacturing of pDNA for gene therapy applications or DNA vaccines.

4.6 References

[1]D. Luo, and W.M. Saltzman, Synthetic DNA delivery systems. Nat Biotechnol 18 (2000) 33-7.

[2]S. Gurunathan, D.M. Klinman, and R.A. Seder, DNA vaccines: immunology, application, and optimization*. Annu Rev Immunol 18 (2000) 927-74.

[3]Wiley, Gene Therapy Clinical Trials Worldwide http://www.wiley.co.uk/genmed/clinical/. (2009).

[4]M.M. Diogo, J.A. Queiroz, and D.M. Prazeres, Chromatography of plasmid DNA. J Chromatogr A 1069 (2005) 3-22.

[5]B. Zaslavsky, Aqueous Two-Phase Partitioning: Physical Chemistry and Bioanalytical Applications, Marcel Dekker, New York, 1995.

[6]P.Å. Albertsson, Partition of cell particles and macromolecules : separation and purification of biomolecules, cell organelles, membranes, and cells in aqueous polymer two-phase systems and their use in biochemical analysis and biotechnology, Wiley, New York, 1986.

[7]H.-K. Rajni, Aqueous Two-Phase Systems: Methods and Protocols (Methods in Biotechnology), Humana Press, Inc, Totowa, New Jersey, 2000.

[8]S.P. Duarte, A.G. Fortes, D.M. Prazeres, and J.C. Marcos, Preparation of plasmid DNA polyplexes from alkaline lysates by a two-step aqueous two-phase extraction process. J Chromatogr A 1164 (2007) 105-12.

[9]H. Barbosa, A.V. Hine, S. Brocchini, N.K. Slater, and J.C. Marcos, Affinity partitioning of plasmid DNA with a zinc finger protein. J Chromatogr A O 1206 (2008) 105-12.

[10]B. Maestro, I. Velasco, I. Castillejo, M. Arevalo-Rodriguez, A. Cebolla, and J.M. Sanz, Affinity partitioning of proteins tagged with choline-binding modules in aqueous two-phase systems. J Chromatogr A 1208 (2008) 189-96.

[11]D. Platis, and N.E. Labrou, Application of a PEG/salt aqueous two-phase partition system for the recovery of monoclonal antibodies from unclarified transgenic tobacco extract. Biotechnol J (2009).

[12]A.C. Roque, C.S. Silva, and M.A. Taipa, Affinity-based methodologies and ligands for antibody purification: advances and perspectives. J Chromatogr A 1160 (2007) 44-55.

[13]Y.H. Kim, J.H. Park, M. Lee, T.G. Park, and S.W. Kim, Polyethylenimine with acid-labile linkages as a biodegradable gene carrier. J Control Release 103 (2005) 209-19.

[14]F.M. Veronese, Peptide and protein PEGylation: a review of problems and solutions. Biomaterials 22 (2001) 405-17.

[15]J.M. Harris, N.E. Martin, and M. Modi, Pegylation: a novel process for modifying pharmacokinetics. Clin Pharmacokinet 40 (2001) 539-51.

[16]M. Zaveckas, A. Zvirbliene, G. Zvirblis, V. Chmieliauskaite, V. Bumelis, and H. Pesliakas, Effect of surface histidine mutations and their number on the partitioning and refolding of recombinant human granulocyte-colony stimulating factor (Cys17Ser) in aqueous two-phase systems containing chelated metal ions. J Chromatogr B Analyt Technol Biomed Life Sci 852 (2007) 409-19.

[17]U. Sivars, J. Abramson, S. Iwata, and F. Tjerneld, Affinity partitioning of a poly(histidine)-tagged integral membrane protein, cytochrome bo3 ubiquinol oxidase, in a detergent--polymer aqueous two-phase system containing metal-chelating polymer. J Chromatogr B Biomed Sci Appl 743 (2000) 307-16.

[18]J. Porath, Immobilized metal ion affinity chromatography. Protein Expr Purif 3 (1992) 263-81.

[19]G. Birkenmeier, M.A. Vijayalakshmi, T. Stigbrand, and G. Kopperschlager, Immobilized metal ion affinity partitioning, a method combining metal-protein interaction and partitioning of proteins in aqueous two-phase systems. J Chromatogr 539 (1991) 267-77.

[20]R.A. Darby, and A.V. Hine, LacI-mediated sequence-specific affinity purification of plasmid DNA for therapeutic applications. FASEB J 19 (2005) 801-3.

[21]R.A. Darby, G.M. Forde, N.K. Slater, and A.V. Hine, Affinity purification of plasmid DNA directly from crude bacterial cell lysates. Biotechnol Bioeng 98 (2007) 1103-8.

[22]H. Pesliakas, V. Zutautas, and B. Baskeviciute, Immobilized metal-ion affinity partitioning of NAD(+)-dependent dehydrogenases in poly(ethylene glycol)-dextran two-phase systems. J Chromatogr A 678 (1994) 25-34.

[23]Z.R. Zhang, M.D. Hughes, L.J. Morgan, A.F. Santos, and A.V. Hine, Fluorescent microplate-based analysis of protein-DNA interactions. II: Immobilized DNA. Biotechniques 35 (2003) 988-90, 992, 994 passim.

[24]S. Sasakawa, and H. Walter, Partition behavior of native proteins in aqueous dextran-poly(ethylene glycol)-phase systems. Biochemistry 11 (1972) 2760-5.

[25]N. Dalloraa, J. Klemza, and P. Filho, Partitioning of model proteins in aqueous two-phase systems containing polyethylene glycol and ammonium carbamate. Biochem. Eng. J. 34 (2007) 92-97.

[26]G.N. Ferreira, G.A. Monteiro, D.M. Prazeres, and J.M. Cabral, Downstream processing of plasmid DNA for gene therapy and DNA vaccine applications. Trends Biotechnol 18 (2000) 380-8.

Chapter 5

Conclusions and Future Perspectives

5.1 Summary of the Central Results

5.1.1 Affinity purification heuristics

The work presented in this thesis constitutes the first experimental study on the affinity partitioning of plasmids in aqueous two-phase systems using protein-based affinity ligands. The results presented provide useful information for the future development and designing of affinity pDNA extraction process using aqueous two-phase systems. In this chapter the major practical implications of the studies discussed in the previous chapters are synthesised. Areas for future research identified in this thesis are also summarised in this section.

On the analysis of the results one has to bear in mind that an ideal system for the affinity extraction of pDNA in aqueous two-phase systems has to comply with a number of rules of thumb or heuristics:

- Remove the most plentiful impurities first
- Specificity of the ligand to the target pDNA without co-purification of impurities
- Simple elution strategy that gives high elution yields
- Stability and robustness of the ligand
- Affinity ligand that can be produce cheaply.

5.1.2 Selection of a suitable ATPS system

ATPS can be divided mainly into two types: polymer–salt, polymer–polymer systems. Polymer-salt systems are not recommended for affinity extraction systems since the salts may interfere on the binding of the affinity ligand to target DNA [1] and thus polymer-polymer systems, such the model PEG/dextran systems, are preferred. For the first time, it was systematically evaluated the partitioning of bacterial cell lysate components in different PEG-dextran systems in order to select systems which accumulates most of the pDNA impurities in

the dextran-rich phase. From the results of presented in Chapter 2, we divided the ATPSs into two main categories, A and B, which differ mainly in the partitioning of total protein from the bacterial cell lysate. In Category A systems, protein accumulated mainly in the bottom dextran phase together with the pDNA and RNA. By contrast, in category B systems protein accumulated in the PEG rich phase, whilst pDNA and RNA accumulated in the bottom phase. The utilisation of affinity ligands that naturally accumulate in the top phase, or that can be modified with PEG molecules (PEGylation) to steer their partitioning to this phase, could be best used in category A systems. The category B system may be suitable for the utilisation of non-modified protein affinity ligands.

Advantages of PEG-dextran systems

The PEG-dextran systems has the following advantages for the utilization of affinity ligands in pDNA production:

- Most of pDNA impurities accumulated in the bottom phase in category A systems (e.g. PEG 600 – DEX 40). A removal of more than 80 % of protein, 99 % of the RNA and ~100% genomic DNA can be achieved.
- In systems composed with PEG 1000 – DEX 500 (category B systems) protein-affinity ligands, such as GST-ZnF and LacI-His_6-GFP fusion proteins, partitioning to the top phase without the need for PEGylation. However, protein based impurities also partition to this phase.
- Unlike polymer-salt systems the partition of bacterial cell lysates components in these polymer-polymer systems varies less from system-to-system composed with different molecular weight polymers.

Disadvantages of PEG- dextran systems

The main disadvantages of PEG-dextran system are:

- Dextran polymer is relative expensive and may limit its application at large-scale.
- Viscosity of some systems may be high (particularly for high molecular weight polymers) that would difficult the operation at bigger scales.
- Proteins impurities, unlike nucleic acids, partitioning more equally between phases in most of the systems.

5.1.3 Affinity Purification of pDNA in ATPS

Two affinity systems were investigated in this thesis for their suitability to purify therapeutic grade plasmid DNA in ATPS:

1) **GST-ZnF/pTS**: a fusion protein, comprised of Glutathione-S-Transferase (GST) and a zinc finger transcription factor (ZnF) that binds to 5´-GGGGCGGCT-3´ sequence within pTS plasmid was utilized to isolate it in ATPS; and

2) **LacI-His6-GFP/pUC19**: a *lac* repressor protein (LacI) fused to a His_6 tag and to Green Fluorescence Protein (GFP) that displays affinity for the *lac* operator sequences ($lacO_1$/$lacO_3$) within pUC19 plasmid.

The tailoring of the affinity protein into the PEG phase was accomplished using two different approaches.

The first method investigated was performed by covalent attachment of the PEG phase-forming polymer to the GST-ZnF affinity protein. In Subchapter 3.1 this approach was initially investigated using the free GST-tag using four different functionalised PEGylation reagents for modification. Two amine selective reagents, 20 kDa mPEG-butyraldehyde (mPEG-ALD) and 20 kDa mPEG-succinimidyl propionate (mPEG-SPA) were evaluated, as were two thiol selective reagents, 5 kDa mPEG-maleimide (mPEG-MAL) and the bis-alkylating 10 kDa mPEG-monosulfone (mPEG-MS). In the same subchapter it was demonstrated that the optimized PEGylation reactions performed with GST

could be translated to the GST fused to a Zinc Finger protein (GST-ZnF) resulting in similar reactions efficiencies and products of PEGylation. In Subchapter 3.2 the affinity partitioning of the plasmid DNA containing the specific recognition sequence for the PEGylated GST-ZnF was studied in more detail.

In Chapter 4 it was tested a second type of PEGylation composed of immobilized metal–ion affinity PEG (PEG-IDA-Cu(II), 10 kDa) which binds to the poly(histidines) tag that is accessible on the surface of LacI-His_6-GFP protein as shown in Figure 1. In both PEGylation methods developed, it was achieved the enhancement on the partitioning of the affinity protein bounded to the target pDNA to the PEG-rich top phase, allowing the extraction of the pDNA from the dextran-rich phase.

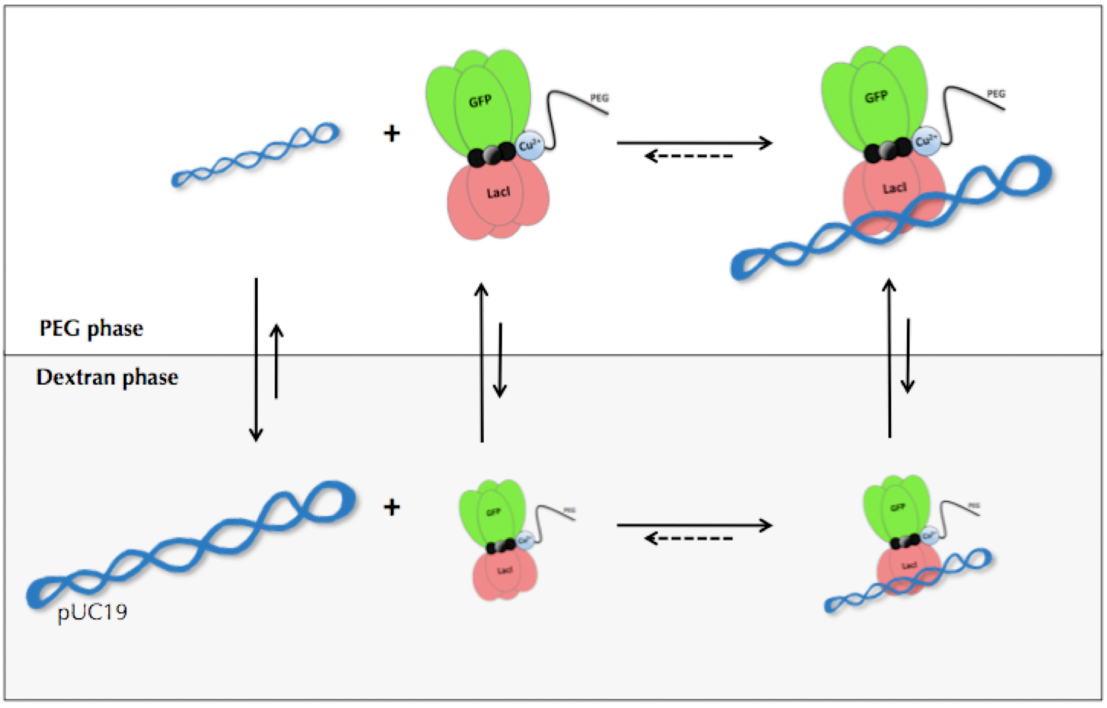


Figure 1 - Scheme of the affinity partition of pUC19 plasmid in ATPS using the LacI-His_6-GFP protein. Sizes of the biomolecules correspond to the relative concentrations in each phase and the arrows indicate the directions of partitioning.

Advantages of using the GST-ZnF/pTS affinity mechanism

The main advantages of using the GST-ZnF ligand approach are:

- Four different PEGylation reactions are now optimized for the PEG conjugation to the free GST-tag which could be translated to other GST fusion proteins
- N-terminal amino acid PEGylation of the GST-ZnF protein could be obtained in the GST-tag domain, which is far from the protein active site within ZnF domain, decreasing the probably of inhibiting binding to the target pDNA.
- Heterogeneous PEGylation products (more than one PEG chains per GST-ZnF) can be used without inhibiting the DNA binding properties. Complicated separations of the different PEGylated species, as result of a PEG-conjugation reaction, are not necessary.
- A 1322-fold increase in the protein partition coefficient (97.5%) after PEGylation is obtained in comparison to the non-PEGylated protein (Kc = 0.013).
- Plasmid DNA partition totally to the top phase in the presence of PEG-GST-ZnF ligand whereas without the ligand partitions 99.9 % to the bottom phase.

Disadvantages of using the GST-ZnF/pTS affinity mechanism

The main disadvantages of the GST-ZnF/pTS affinity approach are as follow:

- The pre-purification of the affinity ligand followed by the PEG conjugation and separation of the non-reacted protein may not be cost effective and relative expensive to produce.
- Thiol reactive PEGylation disrupted the ZnF binding to the target pDNA. The thiol conjugation reactions were much faster (~1h) comparatively to the amine selective mPEG reagents (days). However, the PEG conjugation on the cys site (probably the cys coordinating with zinc ion) disrupted the binding.

- Protein-based affinity ligands, as the GST-ZnF protein, are unstable and easily degraded (GST-ZnF data not shown).

- In practice, a part from the proof-of-concept presented in this thesis, the GST-ZnF would not be a suitable ligand for a commercial pDNA purification process since the pDNA elution from the ligand is impractical. The presence of the ligand may denote a potential safety hazard.

- In addition, a Blast (Basic Local Alignment Search Tool) has shown that both the GST and ZnF moiety of the fusion protein are similar to protein expressed in humans. It has yet to be shown that native protein do not bind to the therapeutic plasmid containing the GST-ZnF recognition sequence.

Advantages of using the PEG-IDA-Cu(II) / LacI-His$_6$-GFP / pUC19 mechanism

The principal advantages of the aforementioned mechanism are:

- Simple mechanism. All the affinity components are added into the bacterial cell lysate and the pDNA is extracted to the top phase.

- Covalent PEGylation of the affinity protein and sequent purifications steps are not necessary. This will decrease the operation costs and processing time as compared to the GST-ZnF affinity approach.

- Very low amounts of the PEG-IDA-Cu(II) ligand (0.14% w/w) is necessary for the partitioning of the affinity protein to the top phase of PEG-dextran systems.

- More than 72% of the plasmid DNA partition to the top phase, without RNA or genomic DNA contamination.

- The pDNA yields obtained in this thesis represents a 9 fold improvement from the chromatographic methods reported previously in literature.
- The ligand can be lyophilized and stored at room temperature and re-hydrated for use when required.
- A Blast search for nearly exact match of LacI domain showed that no protein produced by human genome matches the ligand.

Disadvantages of using the PEG-IDA-Cu(II)/LacI-His$_6$-GFP /pUC19 mechanism

The main limitations of the PEG-IDA-Cu(II)/LacI-His$_6$-GFP /pTS mechanism are:

- Bacterial cell lysate requires to be desalted prior to plasmid extraction. The addition of a chromatographic step for the desalting it will increase the operational costs.
- Relative low pDNA yields (~27%). The dissociation of pDNA from the protein ligand proved to be a challenging task in aqueous two-phase systems.
- Maximum removal of proteins was only about 76 % from the original feed. This is partially due to the protein ligand that stills tightly bounded to the pDNA.
- It maybe possible that the protein ligand binds to other LacO regions in bacterial gDNA. Although, in this work no co-purification of gDNA was observed. This may not be an issue as bacterial strain could be selected or genetically modified in order to produce a genome that contains no lacO sequences.
- pUC19 plasmids employs antibiotic resistance gene. This disadvantage was discussed in Section 1.2.1
- The use of pure protein-based ligands may be relative expensive to produce**.**

5.1.4 Affinity Purification Summary

Two affinity approaches were developed for the affinity extractions of plasmid DNA in ATPS. The PEG-IDA-Cu(II)/LacI-His_6-GFP/pUC19 mechanism is the one that offers higher potential for use in a commercial pDNA purification process. The mechanism is much simpler and faster compared to the GST-ZnF/pTS mechanism and does not require the covalent binding of the PEG phase forming polymer to the affinity ligand previous to the extraction of pDNA. In addition, very small amounts of PEG-IDA-Cu(II) are needed in order to enhance the partition of the ligand bounded to the target pDNA for the phase where less contaminants accumulate.

Nonetheless, for the process to be commercially viable and to comply with the safety guidelines, the elution of pDNA from the ligand has to be improved in the future. It is hypothesized that the elution difficulties are likely to be attributed to a combination of several effects such as the system composition, elution buffer strength or the opposite phase-partitioning of the eluting agent, IPTG, to the LacI/pDNA complex.

When IPTG binds to the affinity protein, the LacI conformation is altered [2; 3] resulting in the release of the pDNA plasmid from the protein. In this study up to 2.5 mM IPTG was unsuccessfully used for pDNA release whereas 1mM concentration in 5 min incubations, was found to be enough for the release of pDNA when the affinity ligand was immobilized into a chromatographic column [3; 4]. When ATPSs are used the first hypothesis to consider in order to explain this phenomenon, is the partition of the IPTG to the opposite phase where the binding complex accumulates, and thus the elution buffer has limited contact with the affinity protein. However, this hypothesis requires further investigation to determine the IPTG partition in these systems.

Alternatively, it is possible that the IPTG-binding site in the LacI fusion protein may not be easily accessible if protein-DNA complex tend to form molecular aggregates originated by the phase forming polymers.

Nonetheless, as the affinity binding is a complex combination of hydrophobic, electrostatic forces and hydrogen binding, the maximum elution yield may only be achieved with the process optimization of several factors at the same time. Further studies have to be planned with this in mind in order to determine which are the optimal conditions for the efficient pDNA dissociation.

Although pDNA elution from LacI binding complex proved to be difficult, the non-optimized method described here, removed more than 75% of the protein yielding proximally 27 % of pUC19 plasmid DNA. This represents about 7.4 µg of pDNA extracted per 1 mL of pUC19 desalted lysate which is about a 9 fold improvement compared to what was previously obtained using LacI-His6-GFP immobilized into a chromatographic support column [4].

The results demonstrate that a high capacity affinity purification process was developed with specificity to target plasmid DNA, without co-purification of structural similar impurities such as RNA or gDNA. The development of such plasmid recovery processes in ATPS will have an impact on the number of unit operations currently in use in the downstream processing of pDNA.

With this dual affinity method, most of other chromatographic unit operations could be removed, except for a buffer exchange stage, in order to removed phase forming components and elution buffers (IPTG/NaCl). This ultimately, will promote the ATPS process from a mainly primary recovery step to a more global operation step, combining the primary and intermediate recovery stages.

Boehringer Ingelheim has determined that more than 50 % of total production costs for pDNA are incurred during the downstream processing [5]. The elimination of some purification steps (e.g. concentration, Ionic-exchange or hydrophobic chromatography) could reduce the downstream purification costs by over 50% and the current total costs for the production of pDNA could be reduced by over 25 % [6]. The affinity aqueous two-phase systems processes are thus foreseeable to have a huge impact on the industrial and economical manufacturing of pharmaceutical grade plasmid DNA in the future.

5.2 Recommendations for Future Work

In order to further elucidate the economical viability and process effectiveness of the affinity approaches developed in this thesis, there are several aspects that require to be addressed before large-scale implementation. In Table 1, the several unanswered questions and areas that deserve further investigation are summarised.

The utilisation of the affinity ligands in polymer-polymer has proved to be suitable, since systems can be selected where all major pDNA impurities accumulates in the bottom phase. However, the industrial application of the model systems used in this study, the PEG-dextran systems, maybe limited due to the high cost of the dextran phase forming polymer (~$400/kg). The use of dextran is only possible when the cost of the product of interest is considerable and therefore compensates the cost of the forming phase polymers. However, inexpensive substitutes of dextran can be considered in future applications, like cellulose, polyvinyl alcohol. Currently, the utilization of a starch derivative Reppal PES, as less expensive bottom phase-forming polymer, is being investigated in our lab [7].

In addition, the use of protein-based affinity ligands in large-scale processes may be a concern owing to the expense of producing the pure protein ligand. A potential less expensive approach will be the utilisation of crude cell lysates of the protein ligands rather than the pure ligand, as used in this work. The economical benefit will be significant since the operation costs associated to the chromatographic purification of the ligand will be removed. The price of PEG-IDA-Cu(II) is believed not to be very significant, as the PEG and the copper metal are low-cost and the immobilized metal-ion ligand is used in very low amounts. Nonetheless, a more detailed economical analysis of the real costs has to be assessed enclosing the overall stages of ligand production.

Table 1 - Summary of topics for further research

Chapter	Limitation / Problem	Topics for further research
Chapter 2 Selection of ATPS for Affinity Purification of plasmid DNA	expensive dextran	Explore cheaper substitutes of dextran, like derivatives of starch, cellulose, polyvinyl alcohol
Subchapter 3.1 Modified Fusion Proteins for ATPS Affinity Applications: PEGylation of GST	thiol-PEGylation disrupt binding activity	Engineer ligands with additional cysteine residues for preferential PEG conjugation
	only one fusion GST-protein tested	Examine the optimized PEGylation reactions with other GST fusion proteins
Subchapter 3.2 Affinity Partitioning of Plasmid DNA with a Zn Protein	pDNA elution	Design alternative ZnF ligands with less affinity for pDNA
		Test chaotropic agents
	expensive covalent PEGylation process	Eliminate chromatographic separation steps for the separation of non-reacted protein from PEGylated protein by using ATPS
		Access the re-utilization/recycling of the ligand
	antibiotic resistance gene	Design vector using antibiotic-free selection such as such as mention in Section 1.2.1
Chapter 4 Dual Affinity Method for Plasmid DNA Purification in Aqueous two-phase systems	pDNA elution	Study alternative second extraction systems with different phase forming-salts such as citrate, ammonium sulfate, etc.
		Design plasmid DNA vector with lower binding constants to LacI affinity protein
		Test natural alternative eluting agents such as such as allolactose
	co-purification of host proteins	Add imidazole (in minimal amounts) to extraction systems
	desalted lysates	Study alternative bacterial cell lysis process without SDS detergent, such as sonication; autolysis bacterial strains, freeze-thaw.
	ligand Stability / re-usability	Synthetic ligands mimicking natural biological recognition
	overall process Cost	Utilization of bacterial lysates of LacI-His_6-GFP ligand rather than pure affinity ligands;
		Continuous mode-operations using ligand recycling

Another aspect that needs to be considered for future research is the evaluation of a continuous-mode extractions processes in ATPS. As the production of such

ligands are expensive, the economical impact for such process will be significantly lower as compared to batch mode process, where a fresh ligand is used in every new extraction. In the continuous-mode operation the cost-effectiveness will be also variable dependent on the number of times the ligands can be re-used. Thus, process optimizations for increase stability and durability of these ligands are also desirable.

Alternatively, the use of synthetic ligands, mimicking natural biological recognition) may result in ligands that are cheaper to produce, more stable and reusable while being exceptional selective to the plasmid DNA. Molecular modelling studies may aid to find an ideal affinity mechanism with optimal binding and elution conditions. This will offer solutions to the above-mentioned shortcomings of affinity extraction of pDNA in ATPS.

Further work must be also focused on the improvement of the elution yields as mentioned above and on the increase of the final pDNA recovery in ATPS. This may require the optimization of the ligand and/or the pDNA sequence in order to maintain specificity but enable easier elution of the bounded pDNA from the affinity ligand. The utilization of other polymer-salt systems (such as PEG-ammonium sulfate or PEG-citrate) as second extraction may be an alternative approach. Testing other elution buffers or systems conditions may also provide high elutions yields. As the IPTG eluting agent have some toxicity issues, pDNA dissociation should be developed with the absence of IPTG, or by replacing it with allolactose, a non-toxic, but more expensive, natural alternative.

To conclude, the ability to combine the process goals of purification, concentration and clarification into one unit operation would revolutionize the production of pharmaceutical grade pDNA. Aqueous two-phase systems meet all the characteristics of an ideal extraction technology since it has the potential to combine all three operations in one single step. However, these systems lack for high selectivity to the target plasmids and consequently are not yet widely adopted at industrial scale.

It is envisaged in this thesis that the knowledge obtained using pDNA affinity

ligands would contribute to development of high selectivity ATPS process, with higher capacity compared to the traditional chromatographic processes. The specificity and biorecognition properties of these ligands for the pDNA capture from a crude feedstock with large and structurally related impurities, such as, high molecular weight RNA, endotoxin and shear genomic DNA, will facilitate the preparation of a pDNA product fulfilling the stringent specifications for market approval.

In addition, as the affinity mechanism relies on the specific recognition sequence, the technology developed here should be ready transferable to any plasmid DNA containing the recognition site, and, potentially, to other forms of nuclei acid such as RNA and oligonucleotides. It is believed that all the improvement features developed in this thesis, will lead to an increase adoption of ATPS for the downstream processing of pharmaceutical-grade plasmid DNA.

5.3 References

[1]S.P. Duarte, A.G. Fortes, D.M. Prazeres, and J.C. Marcos, Preparation of plasmid DNA polyplexes from alkaline lysates by a two-step aqueous two-phase extraction process. J Chromatogr A 1164 (2007) 105-12.

[2]C.E. Bell, and M. Lewis, A closer view of the conformation of the Lac repressor bound to operator. Nat Struct Biol 7 (2000) 209-14.

[3]R.A. Darby, and A.V. Hine, LacI-mediated sequence-specific affinity purification of plasmid DNA for therapeutic applications. FASEB J 19 (2005) 801-3.

[4]R.A. Darby, G.M. Forde, N.K. Slater, and A.V. Hine, Affinity purification of plasmid DNA directly from crude bacterial cell lysates. Biotechnol Bioeng 98 (2007) 1103-8.

[5]Boehringer Ingelheim, Method and device for isolating and purifying a polynucleotide of interest on a manufacturing scale. The Patent Office (UK) (2003) Pat. # WO03051483.

[6]G.M. Forde, Plasmid DNA purification by affinity methods. Doctor of philosophy dissertation (2004) 24.

[7]A. Oliveira, and J.C. Marcos, Plasmid DNA partition and purification in polyethylene glycol /hydroxypropyl starch aqueous two-phase systems. Poster presentation at "Affinity" International Congress. (2009).

"Science never solves a problem without creating ten more"

George Bernard Shaw

Appendices

Appendix A

Protein Quantification in ATPS

Analytical Biochemistry 395 (2009) 108–110

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Notes & Tips

Protein quantification in the presence of poly(ethylene glycol) and dextran using the Bradford method

Helder Barbosa [a], Nigel K.H. Slater [b], João C. Marcos [a,*]

[a] Center of Chemistry, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
[b] Department of Chemical Engineering and Biotechnology, University of Cambridge, Pembroke Street, Cambridge CB2 3RA, UK

ARTICLE INFO

Article history:
Received 30 May 2009
Available online 3 August 2009

ABSTRACT

Some experimental methodologies require the quantification of protein in the presence of polymers like poly(ethylene glycol) (PEG) and dextran (DEX). In the aqueous two-phase system (ATPS) extraction of biomolecules, the interference of these phase-forming polymers on the Bradford quantification assay is commonly recognized. However, how these polymers interfere has not been reported hitherto. In this study we show that while dextran concentrations of 20% (w/w) can be used without error, loss of accuracy occurs for solutions with PEG concentrations >10% (w/w). Above this value a substantial decrease on the assay sensitivity is observed.

Protein determination in the presence of poly(ethylene glycol) (PEG)[1] or dextran (DEX) polymers is commonly required in several applications, namely protein precipitation studies [1], protein grafting (e.g., PEGylation) [2], and in the aqueous two-phase systems (ATPS) extraction of biomolecules [3].

In ATPS there are several methods for protein determination currently in use. The method of Bradford [4] is the most commonly used but the utilization of bicinchoninic acid (BCA) [5] and absorbance at 280 nm (A_{280}) methods have also been reported [3,6,7].

The Bradford method is popular because it uses a single addition of the dye reagent to the sample, it is rapid (~10 min), and it is done at room temperature (RT). In contrast, the BCA assay takes ~30 min to complete and must be performed at 37–60 °C. The A_{280} assay is the only direct and nondestructive method but it requires that no other compounds present in solution absorb at 280 nm.

Polymer–polymer systems have been studied as two-phase systems for the downstream processing of bioproducts such as cells [8], proteins [7], viruses [9], and plasmid DNA [3]. No specific information has been given regarding the interference of polymers in the protein quantitative assay, though interference is generally recognized. To circumvent this problem a series of samples dilutions are usually performed prior to analysis (to diminish the interference of the polymers) and the standards curves are prepared with the same polymer composition as the samples. This could lead to assay errors.

In this study we evaluated the effect of the concentration of the poly(ethylene glycol) and dextran polymers, and their molecular weights, on the Bradford assay sensitivity using the dye commercially available as Coomassie Plus from Pierce (Rockford, IL, USA). The reagent compatibility with temperature, salt type, salt concentration, and pH is usually given in the manufacturer's manual, but no information is given regarding the recommended working concentrations of PEG and DEX polymers.

PEG polymers (St. Louis, MO, USA) of the molecular weight ranging from 600 to 10,000 Da and DEX (GE Healthcare, Chalfont St. Giles) ranging from 40 to 2000 kDa were used with various concentrations in samples containing the protein bovine serum albumin (BSA) (Pierce) as standards. All the quantitative assays were carried out according to the manufacturer's instructions. Briefly, BSA protein samples were prepared in 5 mM Tris-HCl buffer, pH 7.5. Concentrated polymer solutions were added to each sample to attain the required final concentration prior to standardizing the 0.5 mL final volume with water. The Bradford reagent (0.5 mL) was then added and the sample mixed thoroughly. After 10 min incubation at RT the absorbance of each sample was read at 595 nm wavelength (A_{595}). For each concentration three independent measurements were made and the best estimate taken as their mean.

The standards curves of BSA prepared in the presence of increasing concentrations of PEG 600 (%, w/w) are shown in Fig. 1A. For higher PEG concentrations a substantial decrease of the A_{595} was observed for the same protein concentration. For example, for PEG concentrations of 40% (w/w) a decrease of about 80% of the A_{595} values for BSA concentrations of 25 μg/mL was observed. This reduction was only 15% if 10% (w/w) PEG is used. It was also observed that for the higher PEG concentrations a plateau on A_{595} values was reached at lower BSA protein concentrations

* Corresponding author. Fax: +351 253678983.
E-mail address: jcmarcos@quimica.uminho.pt (J.C. Marcos).

[1] *Abbreviations used:* ATPS, aqueous two-phase system; BCA, bicinchoninic acid; BSA, bovine serum albumin; DEX, dextran; PEG, poly(ethylene glycol); RT, room temperature.

0003-2697/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2009.07.045

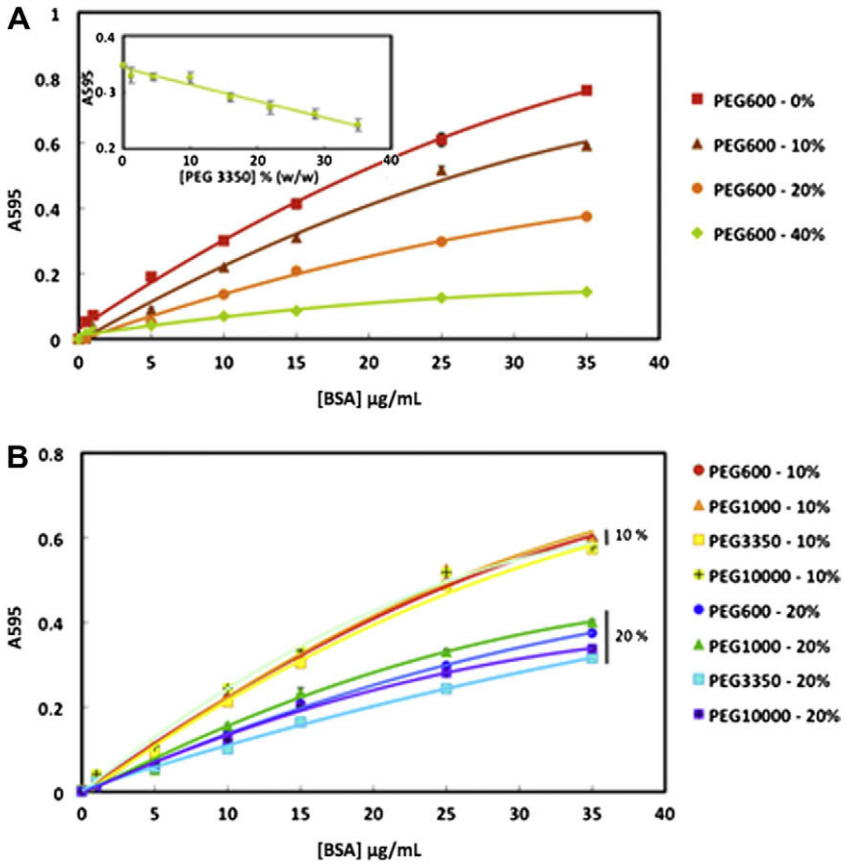


Fig. 1. BSA calibration curves in the presence of PEG using the Bradford reagent. (A) BSA standard curves with increased PEG concentrations (%, w/w) with the MW 600 Da (PEG600). Inset shows the effect of PEG3350 concentration on the A_{595} readings using samples with constant BSA concentration (15 µg/mL). (B) BSA standard curves in the presence of different MW PEG (600; 1000; 3350; 10,000 Da) at 10 and 20% w/w concentrations. Standard deviation (SD) bars are within points.

compared to the low PEG concentration samples. This could lead to the underestimation of real protein concentration in the samples even when calibrations curves are prepared with the same polymer concentrations.

As the inset of Fig. 1A shows, a similar effect was also observed for PEG 3350, indicating that the decrease of A_{595} values for a constant BSA concentration in the samples (15 µg/mL) is dependent on the concentration of the PEG polymers.

Another parameter that is commonly screened when optimizing ATPS for the separation of biological products is the MW of the phase-forming polymers. Differences in size of the polymers can influence the partition behavior of target biomolecules. To evaluate the influence of polymer MW on the protein determination assay we studied PEG solutions at 10% (w/w) and 20% (w/w) as shown in Fig. 1B. For the 10% (w/w) PEG samples, all the standard curves were very similar among the different PEG MWs tested (600; 1000; 3350; 10,000). Therefore the PEG MW at this concentration had no interference with the assay. However, for higher PEG concentrations (20%, w/w) the PEG MW had significant effects on the protein quantification since differences on the A_{595} readings

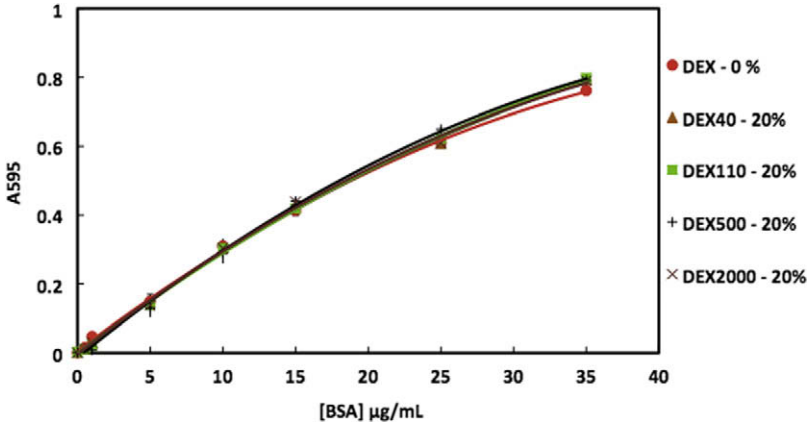


Fig. 2. BSA calibration curves in the presence of DEX using the Bradford reagent. The BSA standard curves with DEX with 40-, 110-, 500-, and 2000-kDa molecular weight and 20% (w/w) concentrations are shown. SD bars are within points.

were easily observed for each sample. At this concentration, the PEG 3350 had a higher interference in the assay compared to the other PEG MWs tested (with PEG 1000 interference the lesser). Nevertheless, all samples at 20% had considerably lower A_{595} values compared to the 10% PEG samples in agreement with Fig. 1A.

In Fig. 2 the effect of dextran concentration and dextran MW in the BSA standard curve preparation was evaluated. In contrast to PEG, dextran concentrations up to 20% (w/w) had the same standard curves as those prepared with 0% dextran. Moreover, it seems that the molecular weights of the dextran tested at 20% (40; 110; 500; 2000 kDa) had no diminishing effects on the A_{595} values compared to the blank standards. Neither the concentration nor the MW of the dextran polymer interfere with the Bradford reagent and much better tolerance is observed compared to the PEG polymer.

Curvature of the protein standard curve can be due to the depletion of free dye at high protein concentrations as well as to protein aggregation for higher protein concentrations [10]. The effect is more apparent when the assay is performed in the presence of PEG, which could be attributed to the fact that PEG can precipitate proteins out of solution. Although no precipitation was observed under the conditions tested a similar mechanism should be present here. This is usually explained on the basis of an excluded volume effect according to which proteins are sterically excluded from the regions of the solvent that are occupied by PEG linear chains [11]. Proteins are thus concentrated and the Coomassie dye has much less volume to bind to the protein with a consequent decrease on the A_{595} readings being observed. It is also well known that different PEG MWs and/or concentrations may cause protein precipitation and this might have an impact on the quantitative determination of protein in solution. These differences are specifically dependent on the protein structure, size, or isoelectric point [11,12]. Since the Coomassie dye may also exhibit some degree of variation response toward different proteins (see manufacturer's manual) it is recommended that for greater assay accuracy the standard curves should be prepared using a pure sample of the target protein to be measured. If that is not possible, then appropriate controls must be made in order to further investigate the influence of PEG on the target protein.

In conclusion, protein determination in PEG/dextran ATPS using Bradford reagent can be greatly influenced by the concentration and molecular weight of PEG. It is recommended that a maximum of 10% (w/w) PEG concentration be used in the samples to give a minimum error in protein quantification in the presence of different PEG MWs. Calibration curves should be made using the same polymer concentration since a decrease of the A_{595} is observed compared to blanks without the polymer and sample dilution. Concentrations of 20% (w/w) of dextran with various MWs can be used for results similar to those of the control blanks.

Acknowledgment

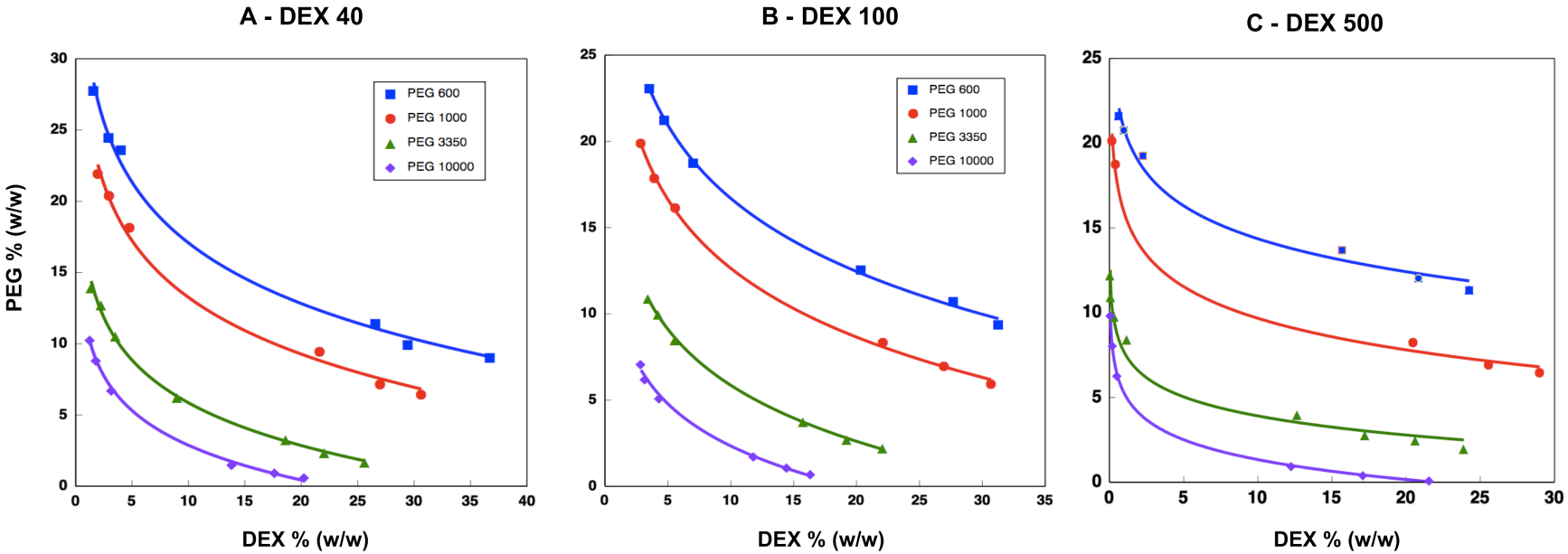
H.B. acknowledges a Ph.D. grant (SFRH/BD/16296/2004/XS49) from FCT, the Portuguese Foundation for Science and Technology

References

[1] S. Raghava, S. Aquil, S. Bhattacharyya, R. Varadarajan, M.N. Gupta, Strategy for purifying maltose binding protein fusion proteins by affinity precipitation, J. Chromatogr. A 1194 (2008) 90–95.
[2] S. Shaunak, A. Godwin, J.W. Choi, S. Balan, E. Pedone, D. Vijayarangam, S. Heidelberger, I. Teo, M. Zloh, S. Brocchini, Site-specific PEGylation of native disulfide bonds in therapeutic proteins, Nat. Chem. Biol. 2 (2006) 312–313.
[3] H. Barbosa, A.V. Hine, S. Brocchini, N.K. Slater, J.C. Marcos, Affinity partitioning of plasmid DNA with a zinc finger protein, J. Chromatogr. A 1206 (2008) 105–112.
[4] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding, Anal. Biochem. 72 (1976) 248–254.
[5] P.K. Smith, R.I. Krohn, G.T. Hermanson, A.K. Mallia, F.H. Gartner, M.D. Provenzano, E.K. Fujimoto, N.M. Goeke, B.J. Olson, D.C. Klenk, Measurement of protein using bicinchoninic acid, Anal. Biochem. 150 (1985) 76–85.
[6] M. Antov, L. Anderson, A. Andersson, F. Tjerneld, H. Stalbrand, Affinity partitioning of a Cellulomonas fimi beta-mannanase with a mannan-binding module in galactomannan/starch aqueous two-phase system, J. Chromatogr. A 1123 (2006) 53–59.
[7] A.M. Azevedo, P.A. Rosa, I.F. Ferreira, M.R. Aires-Barros, Optimisation of aqueous two-phase extraction of human antibodies, J. Biotechnol. 132 (2007) 209–217.
[8] A. Kumar, M. Kamihira, I.Y. Galaev, B. Mattiasson, S. Iijima, Type-specific separation of animal cells in aqueous two-phase systems using antibody conjugates with temperature-sensitive polymers, Biotechnol. Bioeng. 75 (2001) 570–580.
[9] U.G. Vavilova, N.E. Gorev, A.A. Smorodintsev, Concentration of influenza and other respiratory viruses in nasopharyngeal washings by two-phase aqueous polymer systems, Acta Virol. 16 (1972) 83–86.
[10] S.R. Mikkelsen, E. Corto, Bianalytical Chemistry, Wiley, New York, 2004.
[11] D.H. Atha, K.C. Ingham, Mechanism of precipitation of proteins by polyethylene glycols. Analysis in terms of excluded volume, J. Biol. Chem. 256 (1981) 12108–12117.
[12] K.C. Ingham, Precipitation of proteins with polyethylene glycol, Methods Enzymol. 182 (1990) 301–306.

Appendix B

Phase Diagrams

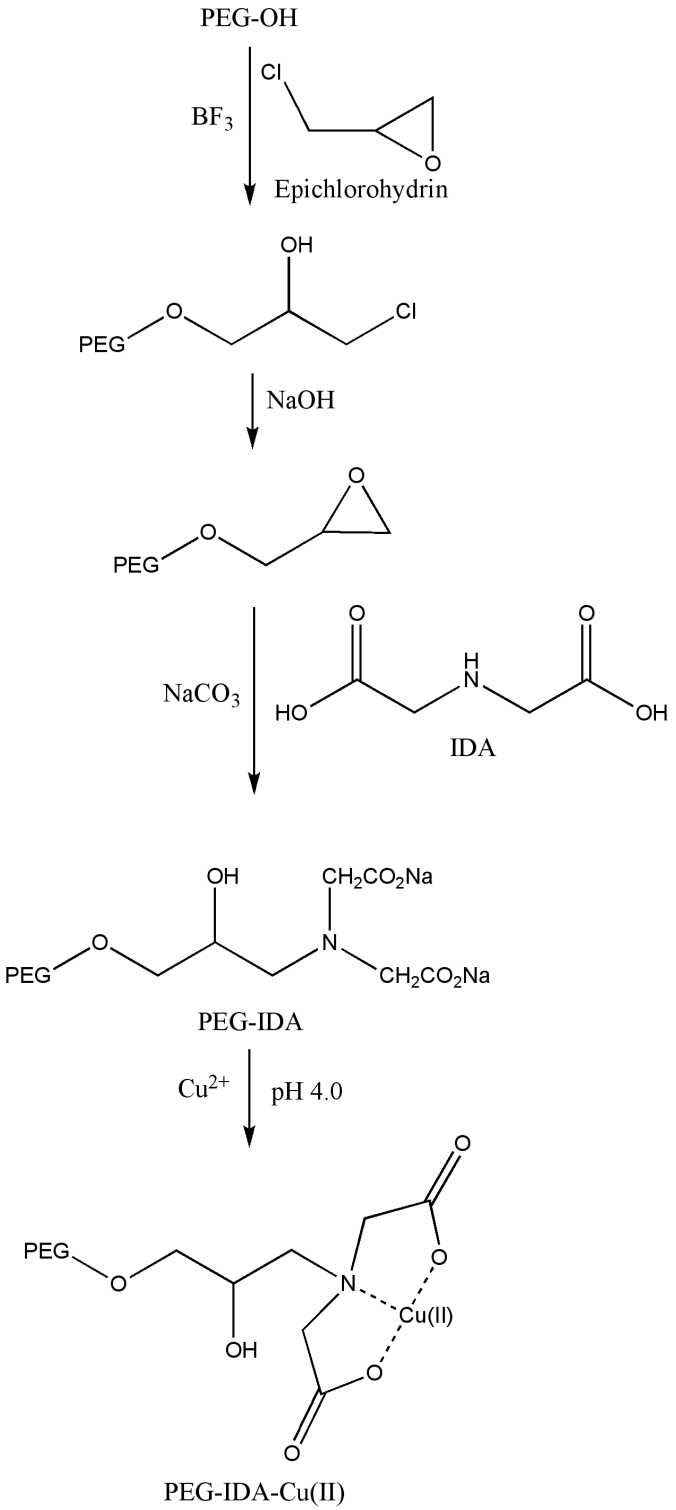


Phase diagrams of each PEG-dextran systems used in this thesis. Binodals using tie-line points. A) The DEX 40 and PEG (600, 1000, 3350, 10,000) phase diagrams. B) DEX 100 and PEG (600, 1000, 3350, 10,000) phase diagrams; C) – DEX 500 and PEG (600, 1000, 3350, 10,000) phase diagrams.

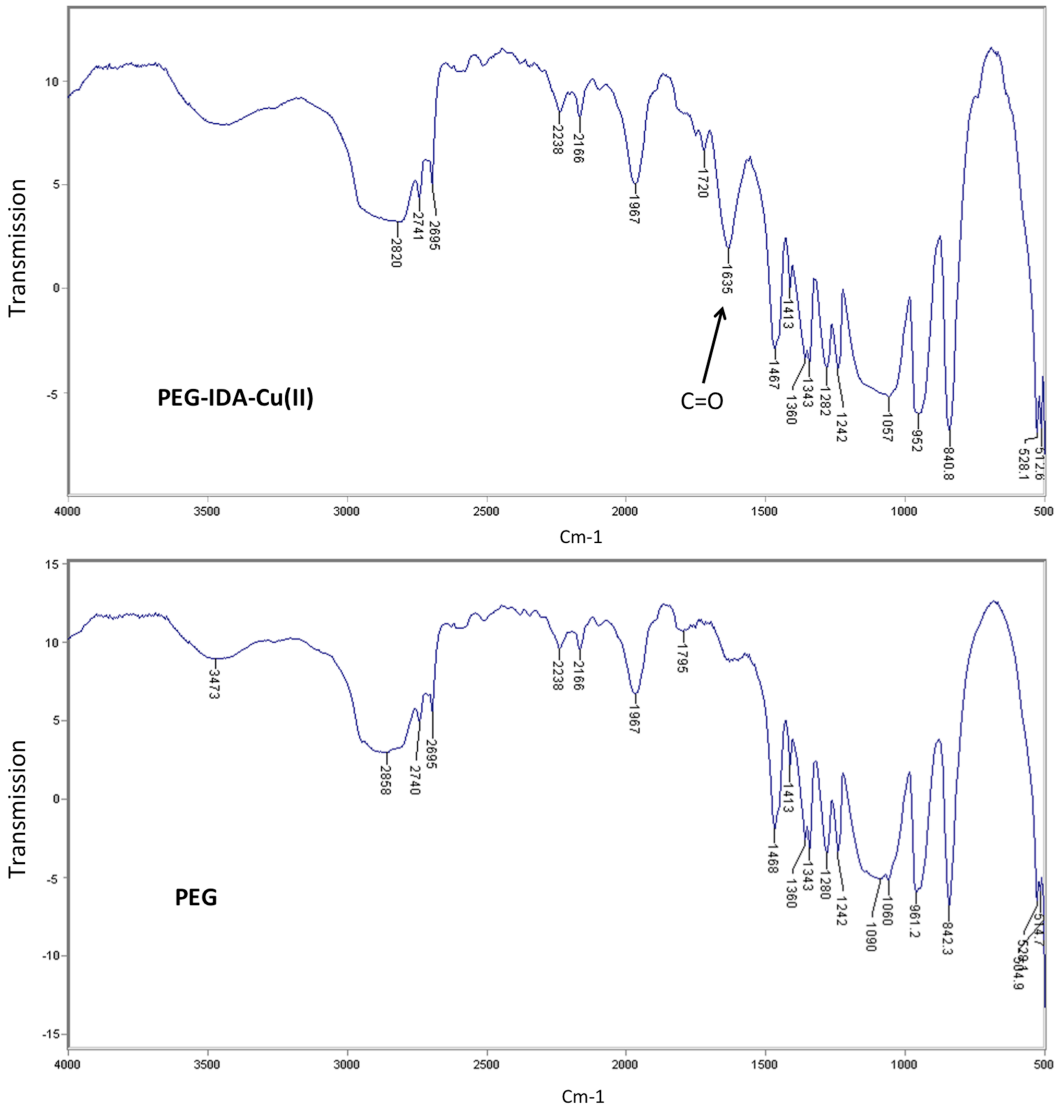
Appendix C

Synthesis of PEG-IDA-Cu(II)

Synthesis route of PEG-IDA-Cu(II)



Fourier Transform Infrared (FTIR) spectra of PEG and PEG-IDA-Cu(II)



Infrared spectra were recorded on a BOMED MB 104 spectrophotometer. IR(KBr): ν (cm^{-1}) 3473 (OH), 1967, 1635 (C=O) 1467.

Atomic Absorption Spectrometry analysis

Standard solutions

Stock solutions of copper metal with a concentration of 1000 ppm (1000 mg/L) were prepared using copper (II) nitrate hexahidrated (Merck, p.a., 99,9 %). An intermediate standard solution of 100 ppm concentration was prepared by dilution of 10.00±0,04 mL of 1000 ppm stock solution in a 100.00 mL volumetric flask and the final volume was set by using diluted nitric acid solution (1:499 mL). Standard solutions used in the calibration procedures have a concentration range from 0.05 to 5.0 ppm and were prepared by dilution of adequate volumes of 100 ppm solution on 100.00±0,04 mL volumetric flask using diluted nitric acid solution (0.02 % v/v) to adjust the final volume.

Samples were prepared by dissolving appropriate amounts of PEG-IDA-Cu(II) in adequate volumes of diluted nitric acid solution (0.02 % v/v).

Flame atomic absoption spectroscopy measurements

The direct aspiration flame atomic absorption (FAA) spectrometry measurements were performed using a Philips-Pye Unicam spectrometer (model SP) using a air/acetylene gas mixture. The wavelength 324.7 nm was used for the determination of metal concentration in the samples.